The discrete wavelet transform as a precursor to leaf area index estimation and species classification using airborne hyperspectral data

Asim Banskota

Dissertation submitted to the faculty of Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Geospatial and Environmental Analysis

Randolph H. Wynne, Chair
Karl F. Huemmrich
James B. Campbell
Ross F. Nelson
Valerie A. Thomas

August 8, 2011
Blacksburg, Virginia

Keywords: *hyperspectral remote sensing, LAI, wavelet transform, DART, LUT inversion*

The discrete wavelet transform as a precursor to leaf area index estimation and species classification using airborne hyperspectral data

Asim Banskota

(ABSTRACT)

The need for an efficient dimensionality reduction technique has remained a critical challenge for effective analysis of hyperspectral data for vegetation applications. Discrete wavelet transform (DWT), through multiresolution analysis, offers oppurtunities both to reduce dimension and convey information at multiple spectral scales. In this study, we investigated the utility of the Haar DWT for AVIRIS hyperspectral data analysis in three different applications (1) classification of three pine species (*Pinus spp.*), (2) estimation of leaf area index (LAI) using an empirically-based model, and (3) estimation of LAI using a physically-based model. For pine species classification, different sets of Haar wavelet features were compared to each other and to calibrated radiance. The Haar coefficients selected by stepwise discriminant analysis provided better classification accuracy (74.2%) than the original radiance (66.7%). For empirically-based LAI estimation, the models using the Haar coefficients explained the most variance in observed LAI for both deciduous plots (cross validation R^2 (CV-R^2) = 0.79 for wavelet features vs. CV-R^2 = 0.69 for spectral bands) and all plots combined (CV-R^2 = 0.71 for wavelet features vs. CV-R^2 = 0.50 for spectral bands). For physically-based LAI estimation, a look-up-table (LUT) was constructed by a radiative transfer model, DART, using a three-stage approach developed in this study. The approach involved comparison between preliminary LUT reflectances and image spectra to find

the optimal set of parameter combinations and input increments. The LUT-based inversion was performed with three different datasets, the original reflectance bands, the full set of the wavelet extracted features, and the two wavelet subsets containing 99.99% and 99.0% of the cumulative energy of the original signal. The energy subset containing 99.99% of the cumulative signal energy provided better estimates of LAI (RMSE = 0.46, R^2 = 0.77) than the original spectral bands (RMSE = 0.69, R^2 = 0.42). This study has demonstrated that the application of the discrete wavelet transform can provide more accurate species discrimination within the same genus than the original hyperspectral bands and can improve the accuracy of LAI estimates from both empirically- and physically-based models.

ACKNOWLEDGEMENTS

Many people provided me with support and help during the completion of this dissertation. Of course, my family tops this list. My parents always dreamt of me pursuing PhD study, and they made all possible efforts on their part to fulfill that dream. My wife, in the midst of her own PhD research, contributed to this thesis with her analytical and technical skills, reviewed and edited my writing, and most importantly, tried to tone down my stress level all the time. My siblings and their family were always there to support me both economically and emotionally.

Thank you Randy for your encouragement, guidance and support, and more importantly, thank you for being such a nice human being. Your professional yet friendly behavior has helped me in the successful completion of this thesis with ease and comfort. You always comforted me by reminding that I was doing great, but at the same time, you never let me lose my focus. I sincerely thank you for giving me freedom to come up with research ideas and providing me with opportunities to work with state-of-the-art remote sensing tools and techniques. It must be very hard for you to advise six PhD students simultaneously, but you did a remarkable job.

My friends have always been central to my life. I would like to thank each one of them. But Prabhat Lamichane stands out for his contribution in helping and guiding me throughout this PhD study. His company was so important to me that I drove to his place in Cincinnatti, Ohio seven times within the span of two years. I also thank Jessica Walker for agreeing to read my thesis and giving her valuable comments and suggestions.

I express my sincere appreciation to my PhD committee members, Dr. Ross Nelson, Dr. James Campbell, Dr. Valerie Thomas, and Dr. Fred Huemmrich for their guidance and constructive comments to improve my thesis. I am also grateful to Dr. Philip Townsend and Shawn Serbin from the University of Wisconsin-Madison, Dr. Patrick Johnson and Bomono Emessiene from Zimmerman Associates, Inc., and Dr. J.A.N. van Aardt, Rochester Institute of Technology for sharing their field and image data for my research. Finally, thanks to Dr. Jean Philippe Gastellu-Etchegorry for providing the DART model for this research and Jeremy Rubio for walking me through the nuts and bolts of the model.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 2: IMPROVING WITHIN-GENUS TREE SPECIES DISCRIMINATION USING THE DISCRETE WAVELET TRANSFORM APPLIED TO AIRBORNE HYPERSPECTRAL DATA

CHAPTER 3: INVESTIGATING THE UTILITY OF THE WAVELET TRANSFORM FOR TEMPERATE FOREST LAI ESTIMATION USING HYPERSPECTRAL DATA

CHAPTER 4: AN INNOVATIVE TECHNIQUE FOR BUILDING A LOOK-UP TABLE FOR EFFICIENT INVERSION OF DART TO ESTIMATE FOREST LAI

CHAPTER 5: INVESTIGATING THE UTILITY OF WAVELET TRANSFORMS FOR INVERTING A 3-D RADIATIVE TRANSFER MODEL USING HYPERSPECTRAL DATA TO RETRIEVE FOREST LAI

CHAPTER 1

GENERAL INTRODUCTION AND OBJECTIVE

1. Introduction

1.1 Remote sensing of forests

Forests are the most widely distributed type of terrestrial vegetation, covering nearly 40% of the global land surface. They play an immense role in regional and global ecosystem processes (Westoby, 1989; Wulder, 1998). Forests store large quantities of carbon in vegetation and soil, exchange carbon with the atmosphere, are sources and sinks of atmospheric carbon, and can be managed to sequester significant quantities of carbon on land (Brown, 1996). Forests are important sites for evaporation and transpiration, and thus are major factors influencing climatic patterns (Danson and Curran, 1993). Trees cycle nutrients through decay process and tree regeneration, thus sustaining important functions for a variety of biological processes in forests (Boring et al., 1981). Healthy forests sustain watershed protection through stabilization of soils and interruption of sediment transport to stream channels (Brooks et al., 1992). Economically, forest products are a major natural resource commodity, valued at billions of dollars worldwide.

Forests can be characterized by three main measures, vegetation structure (e.g., LAI, biomass density, phenology, vegetation condition, etc.), taxonomic composition (e.g., vegetation type, species composition, etc.), and vegetation dynamics (Graetz, 1990). Vegetation dynamics are changes in vegetation structure, and/or taxonomic composition, in time and space. Our ability to detect vegetation dynamics thus depends on our ability to characterize static vegetation structure and taxonomic composition

(Hobbs, 1990). Field based direct measurement of these characteristics can be accurate but is expensive and labor-intensive, and is thus only practical on experimental plots of limited size (Pu et al., 2003). Remote sensing, because of its coverage and repetitiveness, is recognized as a reliable method and a practical means of measuring vegetation attributes.

Leaf area index (LAI) is a basic vegetation structure parameter which controls and moderates different climatic and ecological functions. It is defined as one-half the total green leaf area (all sided) per unit ground surface area (Chen et al., 2002). Since LAI is one of the principal factors controlling canopy reflectance (Asner, 1998), a large body of research has investigated the use of airborne and satellite remote sensing data for its accurate estimation (Combal et al., 2003; Darvishzadeh et al., 2008a; Fassnacht et al., 1997; Gong et al., 1995; Huemmrich et al., 2005; Koetz et al., 2005; Schaepman et al., 2005).The advent of narrow-band hyperspectral remote sensing in recent years has provided new avenues for accurately estimating forest LAI. Similarly, a number of studies have used remote sensing data to identify broad categories of forest taxonomic composition, such as, coniferous versus deciduous stands (Budreski et al., 2007; Chand and Badarinath, 2007; Nelson et al., 1985; van Aardt et al., 2008). The increasing availability and sophistication of remote sensors with high spectral and spatial resolution has offered promise for mapping forests even at the species level (Banskota et al., 2011; Clark et al., 2005; Martin et al., 1998).

1.1.1 Hyperspectral remote sensing for species classification and LAI estimation

The tools for vegetation remote sensing have been developed considerably in the past few decades. Optical remote sensing has expanded from multi-spectral sensors to

imaging spectrometers. Imaging spectrometry, or hyperspectral remote sensing, utilizes sensors that typically have hundreds of narrow, contiguous spectral bands between 400 nm and 2500 nm. Studies have shown that these narrow bands offer significant improvement over broad bands in a wide range of studies including both LAI estimation (Darvishzadeh et al., 2008[a]; Lee et al., 2004; Treitz and Howarth,1999) and species classification (Lucas et al., 2008; Martin et al., 1998; van Aardt et al., 2007).

Estimation of LAI from remote sensing data involves either empirical or physical modeling techniques based on radiative transfer models. In the former, the relationship between LAI and remote sensing data is established through statistical relationships calibrated using small experimental datasets. In the latter, the relationship is based on the physical laws that govern the radiation transfer within vegetation canopies. Previous studies have shown that both techniques can take advantage of the increased information contained in hyperspectral data. Classification of remote sensing data, on the other hand, is carried out primarily using statistical techniques. Hyperspectral remote sensing has made it possible to leapfrog from classifying forests into broad categories (deciduous/coniferous) to species-level classification. However, the luxury of having a large number of narrow bands comes with an additional cost of needing to reduce the dimensionality of the data. The information contained in hyperspectral bands is noisy and highly correlated. The redundancy among bands limits the applicability of using entire bands in either classification or modeling algorithms (for both statistical and physical models). Similarly, the algorithms require that the number of bands should not exceed the number of field observations, and for classifications in particular, the number of field observations should be many times that of the number of remote sensing bands.

Hence, when using hyperspectral data, a common question to be answered is how to reduce the dimensionality and select the right subset of spectral bands for the specific problem.

1.1.2 Band selection versus feature extraction

Selecting bands which contain unique and useful information is a common way of reducing the dimensionality of hyperspectral data (Clark et al., 2005; Schmidt and Skidmore, 2003). The optimal band selection approach is sometimes referred to as feature selection (Yao and Tian, 2003). It involves identification of a subset of the original image bands that best describes a specific application. Instead of having an exhaustive representation of the whole spectrum, it is expected that selecting some key relevant bands can considerably reduce the amount of data without practically losing much information. The relevance of the band is usually identified through some supervised function that ranks the bands according to their significance in predicting a target: eg., species, biophysical variables, etc. However, there is a trade-off among number of bands, relevance and redundancy of bands, and prediction accuracy. It is difficult to have a good balance between these factors and, thus, band selection is a sub-optimal approach to reduce dimensionality (Brunzell and Eriksson, 2000; Miao et al., 2007; Meroni et al., 2004; Weiss et al., 2000).

Alternatively, feature extraction techniques transform the data into a reduced number of dimensions while maintaining most information within the hyperspectral data. The transformation is obtained through some linear combinations of the whole set of original data (Sotoca et al., 2007). Several feature extraction methods are commonly used to reduce the dimensionality of hyperspectral images, including principal

component analysis (Pu and Gong, 2004), singular value decomposition (Lisowski et al., 1996), and maximum noise fraction (Green et al., 1988). A common limitation of feature extraction methods for vegetation applications is that they are based upon the global covariance matrix. Some distinguishable features related to vegetation class and properties are often blurred by the global transforms (Miao et al., 2007).

1.1.3 Wavelet transform

The wavelet transform, a relatively new feature extraction technique, tends to overcome the limitation of traditional techniques by transforming the data such that not only is the dimensionality reduced, but in addition distinguishable features related to vegetation are preserved and enhanced. The wavelet transform can reduce the dimensionality of hyperspectral data by projecting them into a new feature space in which just a few wavelet coefficients represent most of the information in the original data (Blackburn, 2007). These coefficients are related to different scales and wavelength positions of the original data (Hsu, 2007). Scale can be simplistically defined as follows: coarse scale is the big picture, while fine scale shows the details in the data. Thus, going from coarse scale to fine scale, in this context, is equivalent to zooming in. The wavelet transform utilizes variable window sizes to analyze signals at multiple scales (Rioul and Vetterli, 1991). If we looked at a hyperspectral data through a large window, we would notice large features. Similarly, if we looked at the data through a small window, we would notice small features. In general terms, wavelet analysis allows us to see both the forest *and* the trees (Graps, 1995).

The advantages of multi-scale representation of hyperspectral data by wavelet analysis are three-fold. First, the useful information is represented by fewer wavelet

coefficients (Bruce et al., 2002; Peng et al., 2009). It is possible to select only those wavelet coefficients containing useful information and discard others, effectively reducing the dimensionality of the data. Second, the wavelet coefficients exhibit less correlation and reduced noise compared to the original hyperspectral data. Third, wavelets handle low-frequency signals with localized high frequencies very well (Claustres et al., 2004). A hyperspectral signal of vegetation, in particular, is characterized by both subtle variance in reflectance in localized absorption features and gradual variation in the reflectance continuum. Wavelet analysis helps in detecting and resolving such high and low frequency information by zooming in and out at different scales (Hsu, 2007).

The results of wavelet analysis of a signal are wavelet coefficients, which are a function of the scale of the analyzing wavelet and the position of the signal (part of the signal being analyzed). Calculating wavelet coefficients at every possible scale and position is computationally intensive and generates huge amounts of data. In the discrete wavelet transform (DWT), wavelet coefficients are usually sampled at discrete scales and positions. The final results of a DWT decomposition of a spectrum are sets of wavelet coefficients, with each wavelet coefficient directly related to the signal at different positions and scales.

1.1.4 DWT in this study

Research has been reported on the use of wavelet transforms with hyperspectral data for a variety of purposes, including quantifying pigment concentration in vegetation (Blackburn, 2007; Blackburn and Ferwerda, 2008), automatic detection of subpixel hyperspectral targets (Bruce et al., 2001), automatic dimension reduction of

hyperspectral data (Kaewpijit et al., 2003), forest canopy structure identification (Bradshaw and Spies, 1992), forest leaf area index and crown closure mapping (Pu and Gong, 2004), and tropical species identification (Zhang et al., 2006). Results of these studies have shown that wavelet analysis can be an effective tool for reducing the dimensionality of hyperspectral data while maintaining information content for a variety of applications.

Even though DWT has been previously used in classification of remote sensing data, the categorical specificity used for testing the performance of the DWT was not high (i.e., USGS level I or II (Anderson et al., 1976)). No studies have tested the DWT on class information that possesses very similar spectral characteristics (e.g., species-level data within the same genus). Pu and Gong (2004) showed the effectiveness of DWT for LAI estimation using empirical modeling. However, the study lacked a general procedure for selecting the best wavelet coefficients that can potentially provide better estimates of LAI. Very few studies have used dimensionality reduction in physically-based modeling of canopy variables (Darvishzadeh et al., 2008[b]; Meroni et al., 2004). The computationally expensive traditional algorithms used in model inversion might have precluded the use of feature selection/extraction techniques. The increased processing power and recent development of efficient look-up-table (LUT)-based inversion algorithms provide the opportunity to integrate the DWT-based dimensionality reduction techniques in physically-based modeling of LAI.

2. Objectives

The goal of this study was to assess the utility of a discrete wavelet transform as a precursor for improved analysis of hyperspectral data for three remote sensing applications, classification of three pine species (*Pinus spp*.), estimation of LAI using empirical modeling, and estimation of LAI using physically-based modeling. This study examined different sets of wavelet features and their effect on overall classification and LAI retrieval accuracy. A radiative transfer model, Discrete Anisotropic Radiative Transfer (DART), in combination with a LUT-based inversion technique, was used for physically-based estimation of LAI. The main objectives of this study are summarized below. The objectives are discussed in Chapters 2 to 5 of the dissertation as four independent research manuscripts. They are as follows:

- Assess the utility of DWT for classifying three pine species (*Pinus spp*.) using hyperspectral data (Chapter 2)
- Investigate the utility of DWT coefficients for empirically estimating LAI in a mixed temperate forest using hyperspectral data (Chapter 3)
- Develop an efficient LUT building approach to estimate LAI in a temperate deciduous forest by DART inversion using hyperspectral data (Chapter 4)
- Investigate the utility of a discrete wavelet transform for LUT-based inversion of DART to estimate LAI in a temperate deciduous forest (Chapter 5)

References

Anderson, J.R., Hardy, E.E., Roach, J.T., & Witmer, R.E. (1976). A land use and land cover classification system for use with remote sensor data. *U.S. Geological Survey Professional Paper*, 964, 1-41.

Asner, G.P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, 64, 234-253.

Banskota, A., Wynne, R.H., & Kayastha, N. (2011). Improving within-genus tree species discrimination using the discrete wavelet transform applied to airborne hyperspectral data. *International Journal of Remote Sensing*, 32, 3551-3563.

Blackburn, G.A. (2007). Wavelet decomposition of hyperspectral data: a novel approach to quantifying pigment concentrations in vegetation. *International Journal of Remote Sensing*, 28, 2831–2855.

Blackburn, G.A. & Ferwerda, J.G. (2008). Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of Environment*, 112, 1614–1632.

Bradshaw, G.A. & Spies, T.A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *Journal of Ecology*, 80, 205–215.

Boring, L.R., Monk, C.D., & Swank, W.T. (1981). Early regeneration of a clear-cut southern Appalachian forest. *Ecology*, 62, 1244-1253.

Brooks, K.N., Gregersen, H. M., Folliott, P.F., & Tejwani, K.G. (1992). Watershed management: A key to sustainability. In N.P. Sharma (Ed.), *Managing the world's forests* (pp. 455-487). Iowa: Kendall/Hunt.

Brown, S. (1996). Present and potential roles of forests in the global climate change debate. *Unasylva*, 185, 3-9.

Bruce, L.M., Koger, C.H., & Jiang, L. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 2331-2338.

Bruce, L.M., Morgan, C., & Larsen, S. (2001). Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 2217–2226.

Brunzell, H., & Eriksson, J. (2000). Feature reduction for classification of multidimensional data. *Pattern Recognition*, 33, 1741-1748.

Budreski, K.A., Wynne, R.H., Browder, J.O., & Campbell, J.B. (2007). Comparison of segment and pixel-based non-parametric land cover classification in the Brazilian Amazon using multitemporal landsat TM/ETM+ imagery. *Photogrammetric Engineering & Remote Sensing*, 73, 813-827.

Chand, T.R.K., & Badarinath, K.V.S. (2007). Analysis of ENVISAT ASAR data for forest parameter retrieval and forest type classification - a case study over deciduous forests of central India. *International Journal of Remote Sensing*, 28, 4985-4999.

Chen, J.M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S.G., White, H.P., Hall, R.J., Peddle, D.R., King, D.J., Trofymow, J.A., Swift, E., Van der Sanden, J., & Pellikka, P.K.E. (2002). Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sensing of Environment*, 80, 165-184.

Clark, M.L., Roberts, D.A., & Clark, D.B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96, 375–398.

Claustres, L., Boucher, Y., & Paulin, M. (2004). Wavelet-based modeling of spectral bidirectional reflectance distribution function data. *Optical Engineering*, 43, 2327-2339.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., & Wang, L. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, 84, 1-15.

Danson, F.M., & Curran, P.J. (1993). Factors affecting the remotely sensed response of coniferous forest plantations. *Remote Sensing of Environment*, 43, 55-65.

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008)[a]. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in heterogeneous grassland. *Remote Sensing of Environment*, 112, 2592-2604.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F., & Cho, M. (2008)[b]. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63, 409-426.

Fassnacht, K. S., Gower, S. T., MacKenzie, M. D., Nordheim, E. V., & Lillesand, T. M. (1997). Estimating the leaf area index of north central Wisconsin forest using the Landsat Thematic Mapper. *Remote Sensing of Environment*, 61, 229–245.

Gong, P., Pu, R.L., & Miller, J.R. (1995). Coniferous forest leaf-area index estimation along the Oregon transect using compact airborne spectrographic imager data. *Photogrammetric Engineering & Remote Sensing*, 61, 1107-1117 .

Graetz, R.D. (1990). Remote sensing of terrestrial ecosystem structure: an ecologist's pragmatic view. In R.J. Hobbs, & H.A. Mooney (Eds.),*Remote sensing of biosphere functioning* (pp.5- 30). New York: Springer-Verlag.

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering*, 2, 50–61.

Green, A.A., Berman, M., Switzer, P., & Craig, M.D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26, 65-74.

Hobbs, R.J. (1990). Remote sensing of spatial and temporal dynamics of vegetation. *Remote sensing of biosphere functioning*, Ed. Hobbs, R.J., & Mooney, H.A., 5- 30. (New York: Springer-Verlag).

Hsu, P.H. (2007). Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 78-92.

Huemmrich, K.F., Privette, J.L., Mukelabai, M., Myneni, R.B., & Knyazikhin, Y. (2005). Time-series validation of MODIS land biophysical products in a Kalahari woodland, Africa. *International Journal of Remote Sensing,* 26, 4381-4398.

Kaewpijit, S., Le Moigne, J., & El-Ghazawi, T. (2003). Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 863–871.

Koetz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment*, 95, 115-124.

Lee, K.S., Cohen, W.B., Kennedy, R.E., Maiersperger, T.K.,& Gower, S.T. (2004). Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment*, 91, 508–520.

Lisowski, J.J., Cook, C.A., & Shen, S.S. (1996). A SVD method for spectral decomposition and classification of ARES data. *Hyperspectral Remote Sensing and Applications*, *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2281, 14-29.

Lucas, R., Bunting, P., Paterson, M., & Chisholm, L. (2008). Classification of Australian forest communities using aerial photography, CASI and HyMap data. *Remote Sensing of Environment*, 112, 2088-2103.

Martin, M.E., Newman, S.D., Aber, J.D., & Congalton, R.G. (1998). Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment*, 65, 249–254.

Meroni, M., Colombo, R., & Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment*, 92, 195-206.

Miao, X., Gong, P., Pu, R., Raymond, I.C., & Heaton, J.S. (2007). Applying class-based feature extraction approaches for supervised classification of hyperspectral imagery. *Canadian Journal of Remote Sensing*, 33, 162-175.

Nelson, R. F., Latty, R. S., & Mott, G. (1985). Classifying northern forests using Thematic Mapper Simulator data. *Photogrammetric Engineering & Remote Sensing*, 50, 607–617.

Peng, Z.K., Jackson, M.R., Rongong, J.A., Chu, F.L., & Parkin, R.M.(2009). On the energy leakage of discrete wavelet transform. *Mechanical Systems and Signal Processing*, 23, 330-343.

Pu, R., & Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, 91, 212-224.

Pu, R.L., Gong, P., Biging, G.S., & Larrieu, M.R. (2003). Extraction of red edge optical parameters from Hyperion data for estimation of forest leaf area index. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 916-921.

Rioul, O., & Vetterli, M. (1991). Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8, 14–38.

Schaepman, M.E., Koetz, B., Schaepman-Strub, G., & Itten, K.I. (2005). Spectrodirectional remote sensing for the improved estimation of biophysical and -chemical variables: two case studies. *International Journal of Applied Earth Observation and Geoinformation*, 6, 271-282.

Schmidt, K.S., & Skidmore, A. K. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85, 92-108.

Sotoca, J.M., Filiberto, P., & Sanchez, J.S. (2007). Band selection in multispectral images by minimization of dependent information. *IEEE Transactions on Systems, Man, and Cybernetics -Part C: Applications and Reviews*, 37, 258-267.

Treitz, P.M., & Howarth, P.J. (1999). Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystem. *Progress in Physical Geography*, 23, 359-390.

van Aardt, J.A.N., & Wynne, R.H. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: An extension of field-based results. *International Journal of Remote Sensing*, 28, 431-436.

van Aardt, J.A.N., Wynne, R.H., & Scrivani, J.A. (2008). Lidar-based mapping of forest volume and biomass by taxonomic group using structurally homogenous segments. *Photogrammetric Engineering & Remote Sensing*, 74, 1033-1044.

Weiss, M., Baret, F., Myneni, R.B., Pragnere, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20, 3-22.

Wulder, M. (1998). Optical remote sensing techniques for the assessment of forest inventory and biophysical parameters. *Progress in Physical Geography*, 22, 449-476.

Westoby, J. (1989): *Introduction to world forestry*. Oxford: Blackwell

Yao, H.B., & Tian, L. (2003). A genetic-algorithm-based selective principal component analysis (GA-SPCA) method for high-dimensional data feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 1469-1478.

Zhang, J., Rivard, B., Sanchez-Azofeifa, A., & Casto-Esau, K. (2006). Intra- and interclass spectral variability of tropical tree species at La Selva, Costa Rica: implications for species identification using HYDICE imagery. *Remote Sensing of Environment*, 105, 129–141.

CHAPTER 2

IMPROVING WITHIN-GENUS TREE SPECIES DISCRIMINATION USING THE DISCRETE WAVELET TRANSFORM APPLIED TO AIRBORNE HYPERSPECTRAL DATA

Abstract

Discrete wavelet analysis was assessed for its utility in aiding discrimination of three pine species (*Pinus spp.)* using airborne hyperspectral data (AVIRIS). Two different sets of Haar wavelet features were compared to each other and to calibrated radiance, as follows: (1) all combinations of detail and final level approximation coefficients, and (2) wavelet energy features rather than individual coefficients. We applied stepwise discriminant techniques to reduce data dimensionality, followed by discriminant techniques to determine separability. Leave-one-out cross validation was used to measure the classification accuracy. The most accurate (74.2%) classification used all combinations of detail and approximation coefficients, followed by the original radiance (66.7%), and wavelet energy features (55.1%). These results indicate that application of the discrete wavelet transform can improve species discrimination within the *Pinus* genus.

Keywords**:** *AVIRIS, species discrimination, wavelet transform*

Reprint of:

Banskota, A., Wynne, R.H., & Kayastha, N. (2011). Improving within-genus tree species discrimination using the discrete wavelet transform applied to airborne hyperspectral data. *International Journal of Remote Sensing*, 32, 3551-3563.

1. Introduction

Effectively managing forests and assessing forest conditions require accurate mapping of forest species. Applications of remote sensing to forest species mapping over recent years have improved substantially, with advances in spatial, spectral, temporal and radiometric resolutions and more advanced data-processing techniques. Several studies (Clark et al., 2005; Leckie et al., 2005; Martin et al., 1998; van Aardt and Wynne, 2007) have shown the potential of hyperspectral remote sensing data for discriminating among tree species. Increased spectral detail, while clearly enhancing separability, also requires effective data processing techniques to address the increased complexity and dimensionality. Problems arise (e.g., the Hughes phenomenon, Hughes, 1968) if too many spectral bands are used with finite training samples. As the dimensionality increases, the number of training samples required for training a specific classifier increases as well. Hence, dimension reduction has become a significant part of hyperspectral image interpretation. Selecting bands which contain unique information and minimize classification error is a common way of reducing dimension (Clark et al., 2005; Schmidt and Skidmore, 2003). However, band selection often degrades the performance of classifiers by discarding variables with important information (Brunzell and Eriksson, 2000; Miao et al., 2007). Alternatively, feature extraction techniques reduce data dimensionality while maintaining most information (and thus discrimination capability) within an image.

Several feature extraction methods are commonly used to reduce the dimensionality of hyperspectral images, including the principal component analysis (PCA), singular value decomposition (Lisowski and Cook, 1996), and maximum noise

fraction (Green et al., 1988; Lee et al., 1990; Phillips et al., 2009). PCA is based upon the global covariance matrix, and thus is not explicitly sensitive to the class structure of the data (Richards and Jia, 2006). The class structure of the data is also not taken into account with the maximum noise fraction and singular value decomposition. Hence, these techniques contribute little to class separability. Canonical analysis and discriminant analysis feature extraction, while often useful when applied to multispectral data, have some disadvantages when applied to hyperspectral data. These disadvantages include: difficulty in calculating co-variance matrices with high dimensional data, and unreliability of extracted features when the classes have similar means or when a class has a very different mean from other classes (Hsu, 2007; Richards and Jia, 2006). More recently, wavelet transforms have been introduced for analysis of hyperspectral data as an efficient means of feature extraction (Bruce et al., 2002; Moon and Merenyi, 1995). A wavelet is a mathematical function used to divide a function into different frequency components, affording analysis of each component with scale-dependent resolution. A wavelet transform is the representation of a function by wavelets. The strength of the wavelet transform for hyperspectral feature extraction lies in this ability to analyze signal at different resolution or scales (Bruce et al., 2001). The advantages of multi-scale representation of hyperspectral data are twofold. First, subtle variation in spectral features in the original hyperspectral data might be detected at different scales (Hsu, 2007). Second, the useful information is represented by fewer wavelet features, effectively compressing the data (Bruce et al., 2002; Peng et al., 2009).

Research has been reported on the use of wavelet transforms with hyperspectral data for a variety of purposes, including quantifying pigment concentration in vegetation (Blackburn, 2007; Blackburn and Ferwerda, 2008), automatic detection of subpixel hyperspectral targets (Bruce et al., 2001), automatic dimension reduction of hyperspectral data (Kaewpijit et al., 2003), forest canopy structure identification (Bradshaw and Spies, 1992), forest leaf area index and crown closure mapping (Pu and Gong, 2004), and tropical species identification (Zhang et al., 2006). Results of these studies have shown that wavelet analysis can be an effective tool for reducing the dimensionality of hyperspectral data while maintaining information content for a variety of applications.

The goal of this study was to assess the utility of the Haar discrete wavelet transform as a discriminant analysis precursor for classifying three pine species (*Pinus spp.*) using airborne hyperspectral data (AVIRIS). We also examined two different sets of wavelet features and their effect on overall classification accuracy.

2. Wavelet analysis

A brief introduction to wavelet analysis is given in this section. More detailed discussion can be found elsewhere (e.g. Graps, 1995).

An oscillating function with a limited time duration that is irregular, asymmetric, and has an average value of zero can be regarded as a wavelet function. A wavelet transform enables signal (data) analysis at different scales or resolutions by creating a series of shifted and scaled versions of the mother wavelet function. The term “mother” implies that a set of basis functions $\{\psi_{\alpha,b}(\lambda)\}$ can be generated from one main function, or the mother wavelet $\psi(\lambda)$ by the following equation (Bruce et al., 2001),

$$\psi_{a,b}(\lambda) = \frac{1}{\sqrt{a}}\psi\left(\frac{\lambda - b}{a}\right) \quad (1)$$

Where *a* is the scaling factor of a particular basis function and *b* is the translation variable along the function's range.

Wavelet analysis is based upon the idea of projecting a signal onto the basis functions to resolve the signal at different scales or resolutions (Hsu, 2007). Resolving data into different scales is analogous to looking at data with windows of different width. The 'gross' features of data might be visible at coarse scales and 'fine' features with fine scales (Blackburn, 2007). The results of wavelet analysis of a signal are wavelet coefficients, which are a function of the scale of the analyzing wavelet and the position of the signal (part of the signal being analyzed). Calculating wavelet coefficients at every possible scale and position is computationally intensive and generates enormous amounts of data. In the discrete wavelet transform (DWT), wavelet coefficients are usually sampled at some discrete scale and positions based on powers of two (dyadic scales and positions). An efficient way to implement this scheme using filters was developed by Mallat (1989).

The basic idea behind this fast algorithm is to represent the wavelet basis as a set of high-pass and low-pass filters. The high pass and low pass filters are related (their power sum is equal to one) and called quadrature mirror filters. Thus, the first level of DWT decomposition of a signal splits it into a low pass version and a high pass version. The second level of decomposition is performed on the low pass signal obtained from the first level of decomposition. The wavelet decomposition can be iteratively performed

until a maximum scale is reached. The maximum scale is dependent on the signal length and the wavelet basis length. Figure 1 illustrates an example of the dyadic DWT analysis of a hyperspectral signal. The figure shows how the discrete wavelet transform decomposes the original hyperspectral signal into its coarse approximation and detail information. The approximations are the coarse-scale, low-frequency components of the signal. The details are the fine-scale, high-frequency components. At each level *j*, we build the j^{th}-level approximation A_j and a deviation signal called the j^{th}-level detail D_j. Summing up the approximation at level *J* and detail components at all scales, *j* = 1…*J* , *J* being the final decomposition level, results in the original signal, which is illustrated in equation (2) (Blackburn 2007),

$$S = A_J + \sum_{j \le J} D_j \qquad (2)$$

Where *S* is the original signal. This property indicates that the DWT does not lose any information in the original signal during transformation. Computation of a DWT is very fast and the transformed results are often easier to manipulate than the original spectra. The DWT provides a multiscale representation of hyperspectral data with local spectral variation in the spectral bands resolved at different scales.

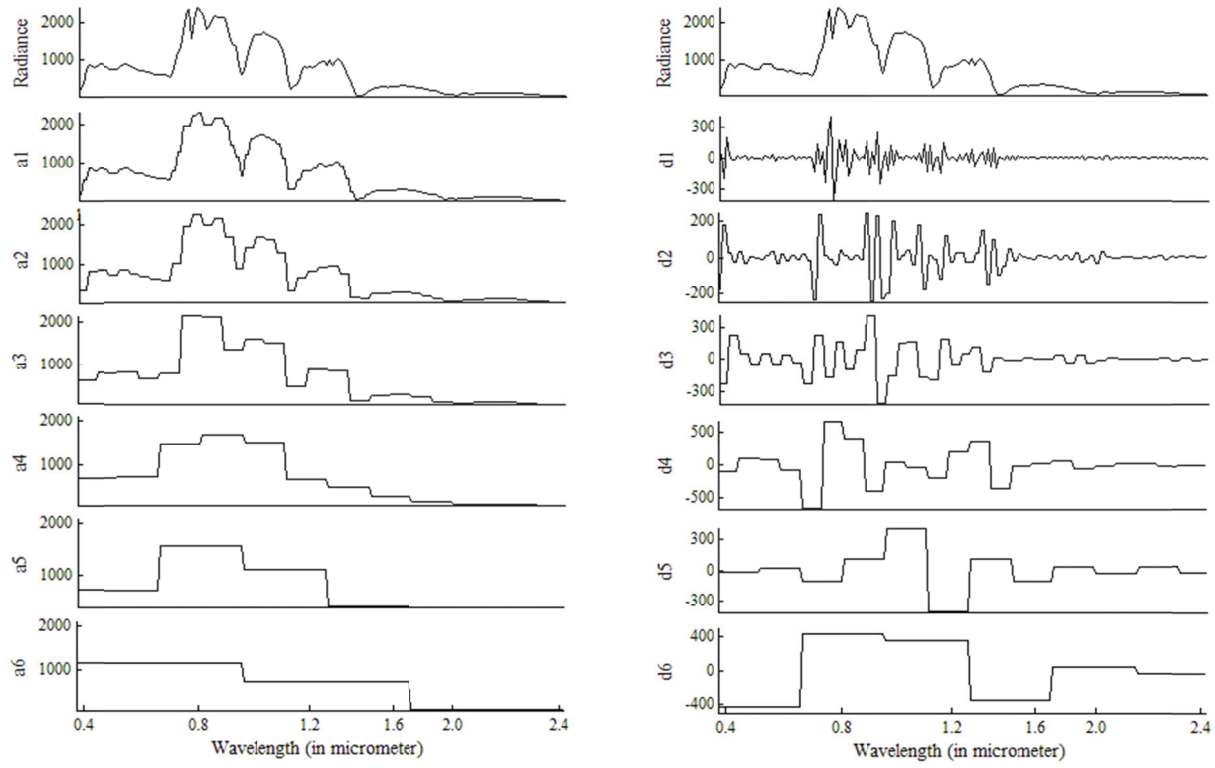

Figure 1: Original hyperspectral radiance (in W m^{-2} sr^{-1}) curve of a randomly chosen shortleaf pine and its approximation (a_1, a_2,…a_6) and details (d_1, d_2…d_6) up to six levels of decomposition.

3. Wavelet feature selection

The final results of a DWT decomposition of a spectrum are sets of wavelet coefficients, with each wavelet coefficient directly related to the amount of energy in the signal at different positions and scales. The next step is to select the wavelet features that contain the maximum amount of useful information for a particular classification problem. The best wavelet features could be a subset of wavelet coefficients, their energy, and any combination of the two (Zhang et al., 2006).

Bruce et al. (2001) decomposed the original hyperspectral signal at the maximum possible level and concatenated all the detail coefficients. Receiver operating

characteristics curves were then used to evaluate the class-discriminating capability of each wavelet coefficient and thus to determine the optimum subset of wavelet coefficients. Rather than using all the coefficients to determine the optimum subset, the wavelet coefficients can also be screened on the basis of their energy. The wavelet transform has good energy concentration and, therefore, most coefficients have very little energy (Peng et al., 2009). Wavelet coefficients with energy values lower than a predetermined threshold value can be discarded without causing significant errors for signal representations (Peng et al., 2009). Okamato et al. (2007) used the retained energy of the coefficient (its square) as a selection criterion. They found that only 50 out of a total of 213 coefficients contained 99.7% of the total energy. They selected the effective variables using a forward stepwise variable selection procedure. Blackburn (2007) and Bruce et al. (2001) used wavelet features based on calculation of an ‘‘energy feature vector’’ from the wavelet coefficients of each spectrum. The 1 × (J + 1) feature vector $\overline{F} = \left[F_j\right]$ was derived from:

$$F_j = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left|W_{j,k}\right|^2} \qquad (3)$$

where K is the number of coefficients at the j^{th} decomposition level, $W_{j,k}$ is the k^{th} coefficient at level j and J is the maximum number of decomposition levels. The feature vector has a length (J + 1) due to the J levels of detail coefficients and one level of final approximation coefficients. The wavelet energy feature vector provides information about how energy of the hyperspectral spectrum is partitioned according to scale (Bruce et al., 2001; Zhang et al., 2006).

4. Methods

4.1 Study area and ground data

The study area (Figure 2) is located in the Appomattox-Buckingham State Forest, Virginia, USA (78°40'30" W, 37°25'9" N). This Virginia piedmont region consists of various pines, upland hardwood, and mixed stands. The three pine species of interest are loblolly pine (*Pinus taeda*), Virginia pine (*Pinus virginiana*), and shortleaf pine (*Pinus echinata*). This study used the same dataset as van Aardt and Wynne (2007). The dataset was collected during September 1999. In situ samples for each of the three pine species were geolocated using a CMT March II GPS unit. Each sampling observation was centered at a stem of the species in question, while at the same time ensuring that the surrounding vegetation (10 m radius) was also of the same species type. All of the GPS data were differentially corrected using CMT PC-GPS software (Version 2.5). The total number of samples was 203, with 89 loblolly, 49 shortleaf, and 65 Virginia pine samples.

4.2 AVIRIS data

The AVIRIS image data were acquired in second week of November 1999 during a low-altitude campaign, resulting in 224 bands, ranging from 400 nm to 2500 nm (approximately 10 nm spectral resolution) at 3.4 m spatial resolution. As reported in van Aardt and Wynne (2007), four flight lines were used in the study, with the first being geo-rectified using identifiable ground control points (GPS collected and differentially corrected), while the remaining three flight lines were registered to an existing ortho-photograph (0.5 m spatial resolution). Root mean square errors were 0.33, 0.23, 0.24, and 0.24 pixels for flight lines 1, 2, 3, and 4, respectively. The four flight lines were

mosaiced using ERDAS Imagine 8.5 software, and subset to the area of interest to facilitate analysis. Raw radiance data were used.

The point map of field sampling plots was overlaid in the AVIRIS image data to extract a full range spectrum of radiance values for each sampling unit. Some of the bands in the shortwave infrared (SWIR) part of the spectrum were noisy (some pixels having either very low or missing values in those bands). Those noisy bands (1353 – 1445 nm, 1803-1961 nm, 2480-2510 nm) were removed, leaving 187 bands for further analysis.

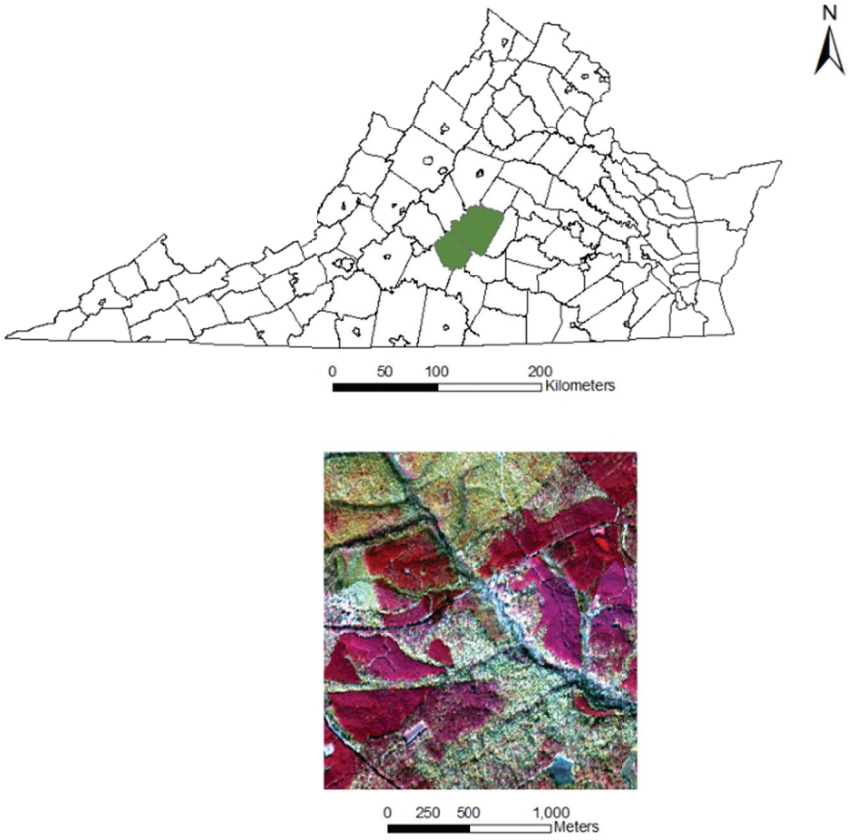

Figure 2: Study area: Commonwealth of Virginia, USA showing general location of Appomattox-Buckingham State Forest (top) and false color composite of AVIRIS image (red, 845 nm; green, 673 nm; blue, 596 nm) (bottom).

4.3 Data analysis

When computing the discrete wavelet transform, two input parameters are required: the choice of mother wavelet and the level of decomposition. We chose the Haar mother wavelet as recent investigations have illustrated its effectiveness for feature extraction of hyperspectral data (Bruce et al., 2001; Li, 2004; Zhang et al., 2006). The value of the decomposition level is determined by the type of wavelet and the original signal length. In this study, the decomposition level was chosen such that it is maximized (7 for 187 bands using the Haar wavelet). The DWT was implemented in Matlab (version 7.4, The Mathworks, Inc.) using a dyadic filter tree as described in section 2. The hyperspectral signal in the spectral domain extracted for each pixel location was passed through a series of low pass and high pass filters.

After feature extraction using the wavelet transform, two sets of wavelet features were selected, namely (1) all combinations of detail and final level approximation coefficients (BEST WAVELET COEFFICIENTS), (2) wavelet energy features rather than individual coefficients (ENERGY FEATURES). While obtaining BEST WAVELET COEFFICIENTS, all the detail wavelet coefficients calculated at each level and approximation coefficients at final level were first concatenated. A simple feature selection procedure with stepwise discriminant analysis (PROC STEPDISC in SAS Version 7.00) was then used to select the optimum set of wavelet features for separating the three different pine species. Stepwise discriminant analysis selects the variables that minimize within statistical group variance while maximizing the between group variance for a given significance (α) level. In this study, the α -level was chosen to be 0.05.

For ENERGY FEATURES, the energy feature vector was calculated using equation 3. The feature vector has length eight (seven levels of detail coefficients and one level of final approximation coefficients). Finally, we had two sets of wavelet input features (BEST WAVELET COEFFICIENTS and ENERGY FEATURES) to use in the classification. To compare the utility of these two sets of wavelet features against the untransformed original radiance data in discriminating among the pine species, a conventional technique of band selection was utilized. Each spectral band in the original hyperspectral signal was evaluated to determine its efficacy for separating known classes. Again, stepwise discriminant analysis was used for this purpose. The best subset of original spectral bands selected by the stepwise procedure was named BEST SPECTRAL BANDS.

4.4 Classification

Classification of the pine species was carried out using three sets of input variables (BEST WAVELET COEFFICIENTS, ENERGY FEATURES and BEST SPECTRAL BANDS) with linear discriminant analysis (PROC DISCRIM in SAS 9.1.3, SAS Institute, Inc.). In linear discriminant analysis (LDA), new variables and values are created. The new variables, called canonical variates, are selected to minimize within group variation and maximize between-group variation (Davies and Fearn, 2008). A two dimensional plot represented by the first two canonical variates provides the best possible view of how well groups are separated (Rencher, 2002). Linear discriminant analysis is sensitive to unequal sample size (Finch and Schneider, 2006; Morrison, 1969). As such, a random subsample of 49 was used for loblolly and Virginia pine to ensure an equal number of sample observations (49) for each class. Classification results were verified

by running a leave-one-out cross-validation routine within the discriminant procedure (with the leave-one-out method, all but one observation is used to apply the classification decision rule and this rule is used to classify the omitted observation). Plots of the canonical variates resulting from the discriminant procedure (RAFisher2cda.m in Matlab 7.4, Trujillo-Ortiz et al., 2004) were used to examine the results visually.

5. Results and discussion

5.1 Feature selection

Stepwise discriminant analysis selected eight wavelet coefficients as the best features for BEST WAVELET COEFFICIENTS. The best detail coefficients correspond to different levels of decomposition (two from scale 1, one from scale 2, two from scale 3, two from scale 4 and one from scale 5). These coefficients from different levels are functions of scale and position (fine detail versus global behavior at various locations in the hyperspectral signal). Selection of coefficients from different levels (or scales) as the best features suggests that both the detail information and the overall trends of the hyperspectral signals were important for discriminating the pine species of interest. A spectral signature curve (Figure 3) was reconstructed (with an inverse DWT) using these eight coefficients to see where the selected features reside in the original spectrum.

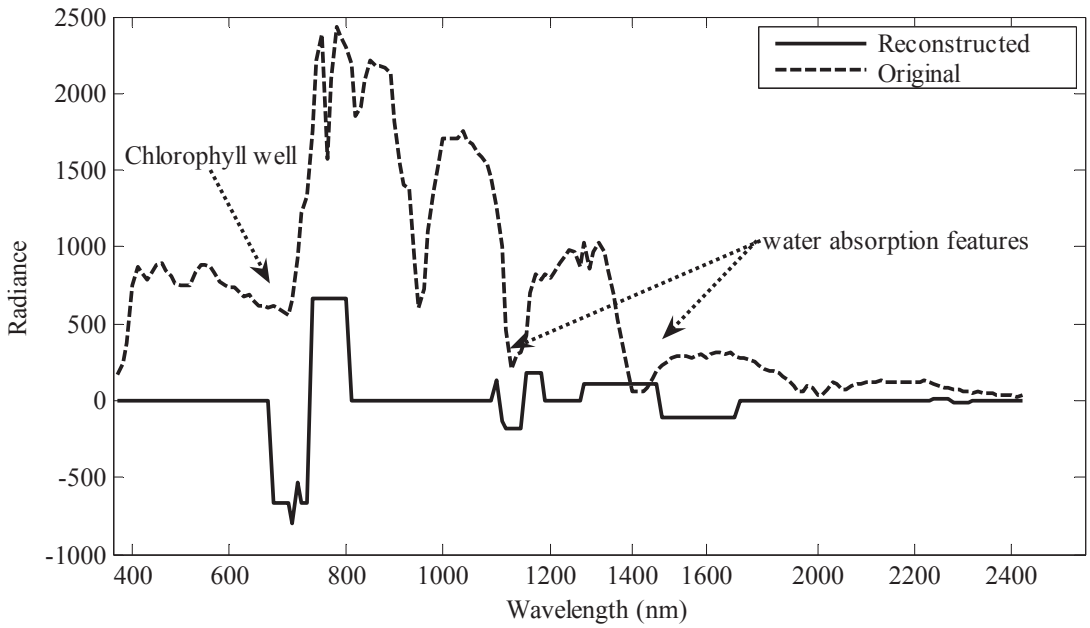

Figure 3: Randomly selected Virginia pine hyperspectral signal and reconstruction of the signal using only the best 8 wavelet detail coefficients selected as BEST WAVELET COEFFICIENTS. Radiance (in W m^{-2} sr^{-1}) is shown on the *y*-axis and wavelength (in nm) on the *x*-axis. Note that the reconstructed radiance curve includes both negative and positive values because only detail coefficients were used to reconstruct the radiance.

Figure 3 shows the reconstructed curve using the best wavelet features and original radiance curve for a randomly chosen Virginia pine spectrum. The reconstruction has a distinct square wave appearance due to the use of the Haar wavelet reconstruction (Bruce et al., 2001). The small spikes in the reconstructed curve arise from the lowest level (finest scale) decomposition; the larger width square wave, in contrast, arises from higher level (coarser scale) decomposition. Accurate matching of the position of wavelet coefficients and the corresponding wavebands in the original spectra is not possible, but is less ambiguous with fine scale coefficients than with coarse scale coefficients. In Figure 3, note that the best wavelet coefficients are centered around the chlorophyll well, red-edge, and the infrared region of the spectrum. One of the best detail coefficients was located at ≈ 700 nm and other one was located at ≈ 1075 nm. The

coarser scale coefficients were also located around the red edge and a water absorption region (≈ 1400 nm).

Using a similar stepwise procedure with the original radiance data, five wavebands centered on 547 nm, 596 nm, 673 nm, 711 nm and 930 nm were chosen as the best features for BEST SPECTRAL BANDS at the 0.05 significance level. Visible and near-infrared regions were important in the spectral separation of the three pine species. However, in contrast to the wavelet features, the water absorption regions were not as important.

5.2 Classification

Cross validation results for the classifications are tabulated in Table 1. Classification with BEST WAVELET COEFFICIENTS provided the highest overall cross-validation accuracy (74.2%). Classification with the BEST SPECTRAL BANDS (66.7%) was more accurate than classification using the wavelet ENERGY FEATURES (55.1%).

Table 1: Cross validation results (percent accuracy) from discriminant analyses for the three different input datasets.

Input features	Loblolly (%)	Shortleaf (%)	Virginia (%)	Overall (%)
BEST SPECTRAL BANDS	61.2	53.1	83.7	66.7
BEST WAVELET COEFFICIENTS	87.8	51.0	83.7	74.2
ENERGY FEATURES	51.0	44.9	69.4	55.1

The results of this study show that Haar discrete wavelet analysis is a useful feature extraction tool for classifying the three pine species using hyperspectral data. DWT not only helped to summarize the important information in hyperspectral data at different scales, but also enhanced the subtle spectral variation around key absorption features. The importance of the red edge and visible part of the spectrum for species

discrimination has been demonstrated in many studies (Kokaly et al., 2003; Schmidt and Skidmore, 2003; van Aardt and Wynne, 2007). Similarly, varying spectral response around water absorption features for different species has been found to be equally useful in species discrimination (Clark et al., 2005; Pu, 2009; Stimson et al., 2005). BEST WAVELET COEFFICIENTS contained detail coefficients from the red edge and visible regions at different scales, as well as coefficients concentrated around water absorption features that afforded better separation among classes.

The accuracy of the classification varies with the choice of wavelet feature set as input to the classification. The best coefficients stemmed from fine scale (scale = 1 and 2) as well as coarse scale (scale = 3, 4, and 5) decompositions in WAVELET COEFFICIENTS. Hence, this feature set provided information on both narrow and wide absorption features. Similar results were found by Koger et al. (2003) while detecting pitted morning glory in soybean. Bruce et al. (2002) also found both fine scale and coarse scale coefficients important for detecting weeds in agricultural crops. With the ENERGY FEATURES, cross validation accuracy was greatly decreased. The relative inability of ENERGY FEATURES to enable separation of the pine species could be due in part to the loss of information from important portions of the vegetation spectra (such as the red edge, chlorophyll well, and water absorption features). Contrary to our results, Zhang et al. (2006) found that the BEST WAVELET COEFFICIENTS and ENERGY FEATURES performed equally well for enhancing spectral separability of tropical tree species at La Selva, Costa Rica (though only 17 trees were sampled). We posit that the difference between our results and those of Zhang et al. (2006) arises from the much more limited variability among the analyzed spectra in our study, since

they were all from within a single (*Pinus*) genus versus five species from five different genera in the study by Zhang et al. (2006). However, other differences between the studies could have been a factor, including differences in the characteristics of the data from the airborne sensors (AVIRIS vs. HYDICE) and sampling (single pixels vs. multipixel averages).

Figures 4(a), and 4(b) show the canonical discriminant plots for the classification using BEST WAVELET COEFFICIENTS and BEST SPECTRAL BANDS, respectively. The group separability, or lack thereof, among species can be visually examined in these canonical plots. Separation among the three groups is possible, although there is still some confusion among classes. Figures 4(a), and 4(b) show that the best separation of loblolly pine from the other two species is achieved along the first canonical variate, and the best separation between shortleaf and Virginia pine is achieved along the second canonical variate. The loblolly pine species is nicely grouped in Figure 4(a), whereas there is more intermixture of loblolly pine with other two species in Figure 4(b). This visual interpretation can be confirmed with the results in Table 1. Cross validation accuracy of loblolly pine was greatly increased from 61.2% with BEST SPECTRAL BANDS to 87.8% with BEST WAVELET COEFFICIENTS. Figure 4(a) shows that loblolly pine is separable whereas shortleaf pine is mixed with Virginia pine. Confusion between shortleaf and Virginia pine could be due to the similarity of their foliage structure, as well as their both having shorter, visibly darker needles than loblolly pine (van Aardt and Wynne, 2007). However, the confusion between two classes seems to impact primarily the error rate for shortleaf classification, as shortleaf has the lowest cross validation accuracy (see Table 1). In this study, we did not further examine the

cause of the greater misclassification rate for shortleaf pine, as our main objective was to investigate the utility of DWT for feature extraction.

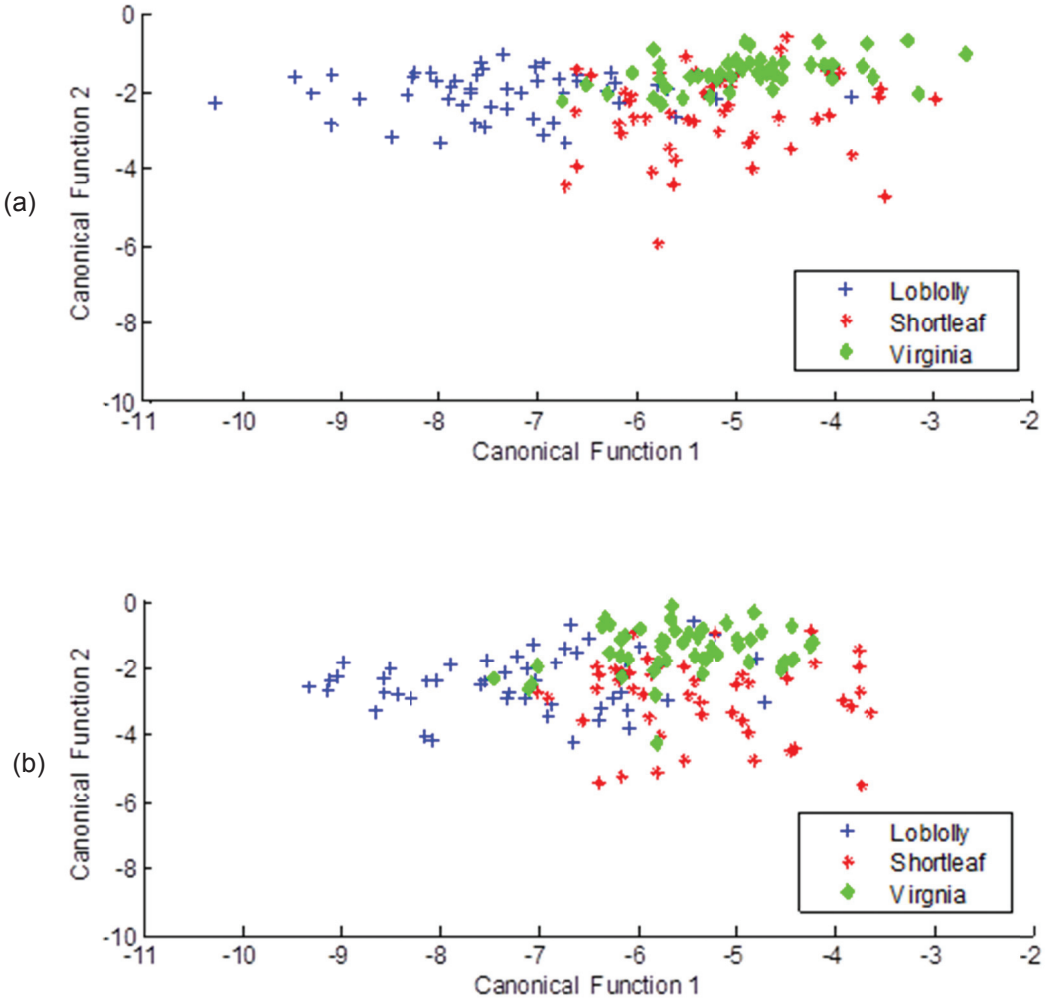

Figure 4: Canonical plots for two different datasets: (*a*) BEST WAVELET COEFFICIENTS, and *(b)* BEST SPECTRAL BANDS.

6. Conclusion

The goal of this study was to assess the utility of the Haar discrete wavelet transform as a discriminant analysis precursor for classifying three pine species (*Pinus spp.*) using airborne hyperspectral data (AVIRIS). Two sets of wavelet features were

examined for their effect on overall classification accuracy. Our results suggest that the Haar discrete wavelet transform can be a very useful tool for pine species identification. Using all combinations of wavelet coefficients with stepwise linear discriminant analysis for variable selection resulted in the highest accuracy. The results of the stepwise selection of wavelet coefficients showed that both fine-and-coarse scale detail coefficients related to key absorption features contained important information for classification of pine species. The lowest separability among pine species was observed when wavelet energy feature vectors were used as inputs to the classification, reiterating the need for both fine- and coarse-scale detail coefficients when discriminating among pine species.

References

Blackburn, G.A. (2007). Wavelet decomposition of hyperspectral data: a novel approach to quantifying pigment concentrations in vegetation. *International Journal of Remote Sensing*, 28, 2831–2855.

Blackburn, G.A., & Ferwerda, J.G. (2008). Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of Environment*, 112, 1614–1632.

Bradshaw, G.A., & Spies, T.A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *Journal of Ecology*, 80, 205–215.

Bruce, L.M., Koger, C.H., & Jiang, L. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 2331–2338.

Bruce, L.M., Morgan, C., & Larsen, S. (2001). Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 2217–2226.

Brunzell, H., & Eriksson, J. (2000). Feature reduction for classification of multidimensional data. *Pattern Recognition*, 33, 1741–1748.

Clark, M.L., Roberts, D.A., & Clark, D.B. (2005). Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. *Remote Sensing of Environment*, 96, 375–398.

Davies, A.M.C., & Fearn, T. (2008). Back to basics: multivariate qualitative analysis, canonical variates analysis. *Spectroscopy Europe*, 20, 18–20.

Finch, W.H., & Schneider, M.K. (2006). Misclassification rates for four methods of group classification. *Educational and Psychological Measurement*, 66, 240–257.

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering*, 2, 50–61.

Green, A.A., Berman, M., Switzer, P., & Craig, M.D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26, 65–74.

Hsu, P.H. (2007). Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 78-92.

Hughes, G.F. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14, 55–63.

Kaewpijit, S., Le Moigne, J., & El-Ghazawi, T. (2003). Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 863–871.

Koger, C.H., Bruce, L.M., Shaw, D.R., & Reddy, K.N. (2003). Wavelet analysis of hyperspectral reflectance data for detecting pitted morning glory (*Ipomoea lacunosa*) in soybean (*Glycine max*). *Remote Sensing of Environment*, 86, 108–119.

Kokaly, R.F., Despain, D.G., Clark, R.N., & Livo, K.E. (2003). Mapping vegetation in Yellowstone National Park using spectral feature analysis of AVIRIS data. *Remote Sensing of Environment*, 84, 437–456.

Leckie, D.G., Tinis, S., Nelson, T., Burnett, C., Gougeon, F.A., Cloney, E., & Paradine, D. (2005). Issues in species classification of trees in old growth conifer stands. *Canadian Journal of Remote Sensing*, 31, 175–190.

Lee, J.B., Woodyatt, A.S., & Berman, M. (1990). Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform. *IEEE Transactions on Geoscience and Remote Sensing*, 28, 295–304.

Li, J. (2004). Wavelet-based feature extraction for improved endmember abundance estimation in linear unmixing of hyperspectral signals. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 644–649.

Lisowski, J.J., Cook, C.A., & Shen, S.S. (1996). A SVD method for spectral decomposition and classification of ARES data. *Hyperspectral Remote Sensing and Applications, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2281, 14–29.

Mallat, S. (1989). A theory for multi-resolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674–693.

Martin, M.E., Newman, S.D., Aber, J.D., & Congalton, R.G. (1998). Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment*, 65, 249–254.

Miao, X., Gong, P., Pu, R., Raymond, I.C., & Heaton, J.S. (2007). Applying class-based feature extraction approaches for supervised classification of hyperspectral imagery. *Canadian Journal of Remote Sensing*, 33,162–175.

Moon, T., & Merenyi, E. (1995). Classification of hyperspectral images using wavelet transforms and neural networks. *Wavelet Applications in Signal and Image Processing III, Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2569, 725–735.

Morrison, D.G. (1969). On the interpretation of discriminant analysis. *Journal of Marketing Research*, 6, 156–163.

Okamoto, H., Murata, M., Kataoka, T., & Hata, S. (2007). Plant classification for weed detection using hyperspectral imaging with wavelet analysis. *Weed Biology and Management*, 7, 31–37.

Peng, Z.K., Jackson, M.R., Rongong, J.A., Chu, F.L., & Parkin, R.M. (2009). On the energy leakage of discrete wavelet transform. *Mechanical Systems and Signal Processing*, 23, 330–343.

Phillips, R.D., Blinn, C.E., Watson, L.T., & Wynne, R.H. (2009). An adaptive noise filtering algorithm for AVIRIS data with implications for classification accuracy. *IEEE Transactions on Geoscience and Remote Sensing*, 47, 3168–3179.

Pu, R., & Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, 91, 212–224.

Pu, R.L. (2009). Broadleaf species recognition with in situ hyperspectral data. *International Journal of Remote Sensing*, 30, 2759–2779.

Rencher, C.A. (2002). *Methods of Multivariate Analyses.* (2nd ed.). New York: Wiley.

Richards, J.A., & Jia, X. (2006). *Remote Sensing Digital Image Analysis*. (4th ed.). Berlin: Springer.

Schmidt, K.S., & Skidmore, A.K. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85, 92–108.

Stimson, H.C., Breshears, D.D., Ustin, S.L., & Kefauver, S.C. (2005). Spectral sensing of foliar water conditions in two co-occurring conifer species: *Pinus edulis* and *Juniperus monosperma*. *Remote Sensing of Environment*, 96, 108–118.

Trujillo-Ortiz, A., Hernandez-Walls, R., & Perez-Osuna, S. (2004). RAFisher2cda: canonical discriminant analysis. A MATLAB file. Available online at: http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=4836 (accessed 3 March 2009).

van Aardt, J.A.N., & Wynne, R.H. (2007). Examining pine spectral separability using hyperspectral data from an airborne sensor: an extension of field-based results. *International Journal of Remote Sensing*, 28,431–436.

Zhang, J., Rivard, B., Sanchez-Azofeifa, A., & Casto-Esau, K. (2006). Intra- and interclass spectral variability of tropical tree species at La Selva, Costa Rica: implications for species identification using HYDICE imagery. *Remote Sensing of Environment*, 105, 129–141.

CHAPTER 3

INVESTIGATING THE UTILITY OF THE WAVELET TRANSFORM FOR TEMPERATE FOREST LAI ESTIMATION USING HYPERSPECTRAL DATA

Abstract

The objective of this study was to determine whether the Haar discrete wavelet transform, when applied to reflectance spectra obtained from hyperspectral data, can improve empirical estimation of LAI in northern temperate forests. The study area comprised a range of deciduous, coniferous, and mixed forest types across several ecoregions within the state of Wisconsin. LAI was estimated for 32 ground plots (17 deciduous, 4 mixed and 11 coniferous), each covering a 60 X 60 m area, using digital hemispherical photography. Plot spectra were extracted from radiometrically and atmospherically corrected AVIRIS data and transformed into wavelet features using the Haar wavelet. Separately, subsets of spectral bands and the Haar wavelet features selected by a genetic algorithm were used as independent variables in multiple linear regressions. The dependent variable was *in situ* LAI observations from either (1) the 17 deciduous plots or (b) all 32 plots independent of forest type. Leave-one-out cross validation was used to measure prediction accuracy. The models using wavelet coefficients explained the most variance for both deciduous plots (CV-R^2 = 0.79, CV-RMSE = 0.31 for wavelet features vs. CV-R^2 = 0.69, CV-RMSE = 0.44 for spectral bands) and all plots combined (CV-R^2 = 0.71, CV-RMSE = 0.46 for wavelet features vs. CV-R^2 = 0.50, CV-RMSE = 0.60 for spectral bands). However, the forest-type specific models were better than the models using all plots combined. Overall, multiscale

wavelet features appear superior to band reflectances for estimating temperate forest LAI using hyperspectral data.

Keywords: *AVIRIS, LAI, wavelet, genetic algorithm*

1. Introduction

Leaf area index (LAI) is a basic vegetation canopy property which controls and moderates different climatic and ecological functions (Chen et al., 2002; Leblanc and Chen, 2001; Huemmrich. et al., 2005; Malinowski et al., 2008; Meroni et al., 2004; Myneni et al., 2002). It controls light interception and thereby CO_2 fixation, canopy photosynthesis, and stand productivity (Gower, 2003; Gower et al., 2001; Turner et al., 2003). It affects hydrological processes and litter production and thus the dynamics of soil water and nutrient cycling (Oren et al., 1998). Thus, most ecosystem process models that simulate carbon and hydrologic cycles require LAI as an input variable (Gower et al., 1999). Since LAI is one of the principal factors controlling canopy reflectance (Asner, 1998), a large body of research has investigated the use of airborne and satellite remote sensing data for its accurate estimation (Cohen et al., 2003; Fassnacht et al., 1997; Gong et al., 1995; Hansen and Schjoerring, 2003; Huemmrich et al., 2005; Koetz et al., 2005; Schaepman et al., 2005). The most widely used approach is to establish an empirical relationship between LAI measured *in situ* and spectral vegetation indices (SVIs) calculated from spectral reflectance in two or three bands (Haboudane et al., 2004). The major shortcoming with empirical approaches is that the relationship saturates at dense canopy conditions characterized by high LAI (Broge and Leblanc, 2000; Chen et al., 2002). The other limitation is that SVIs are sensitive to many

different factors apart from variation in LAI such as variation in leaf optical properties and background spectral reflectance (Goward et al., 1994).

Hyperspectral sensors enable measurement of surface reflectance in narrow spectral bands, providing a capability to analyze canopy by absorption features and over a near continuous spectrum (Asner, 1998; Blackburn, 2007; Curran, 1994; Schmidt and Skidmore, 2003). Both the absorption features and overall shape of the reflectance curve have been found to be sensitive to variability in LAI (Asner, 1998). Darvishzadeh et al. (2008) and Lee et al. (2004) found that the relationship between measured and estimated LAI can be better explained by multiple regression using a combination of narrow bands from hyperspectral data than univariate methods using narrow band SVI. One of the major caveats when analyzing hyperspectral remote sensing data is the greater noise and correlation among spectral bands. Statistical models can suffer from multi-collinearity (Geladi and Kowalski, 1986) and overfitting (Coops et al., 2003; Thenkabail et al., 2000) when a large number of redundant bands are used as predictive variables. Hence, effective use of hyperspectral data for empirical LAI estimation requires dimensionality reduction. However, dimensionality reduction also leads to the loss of some useful features offered by hyperspectral data, such as information about the overall shape of a reflectance continuum, gradual and abrupt slope changes between neighboring bands, etc.

In light of the above discussion, the wavelet transform, a signal processing technique, has become increasingly important to numerous vegetation-related applications of hyperspectral remote sensing (Banskota et al., 2011; Blackburn, 2007; Blackburn and Ferwerda, 2008; Bradshaw and Spies, 1992; Bruce et al., 2001; Cheng

et al., 2011; Kaewpijit et al., 2003; Liu et al., 2011; Pu and Gong, 2004; Zhang et al., 2006). The wavelet transform reduces the dimensionality of hyperspectral data by projecting them into a new feature space in which just a few wavelet coefficients represent most of the information in the original data. Wavelet representation of hyperspectral data also conveys additional information such as the location and nature of high frequency features (narrow absorption features, red-edge inflection point, and noisy bands), and the magnitude and shape of the reflectance continuum at different scales and positions (Banskota et al., 2011).

Pu and Gong (2004) compared the wavelet 'energy features' to the original spectral bands and principal components of Hyperion data for LAI estimation. The energy features were calculated as root mean square of the wavelet coefficients at different scales. The energy features represents how the energy (equivalent to sum of squares of reflectance) is partitioned across wavelet coefficients at multiple scales (Bruce et al., 2001; Zhang et al., 2006). However, the features do not provide any measure of the energy distribution across different wavelength positions. For vegetation applications in particular, the latter is more critical because the coefficients related to specific wavelength regions of hyperspectral data resolved at different scales might be more useful than coefficients related to other regions and scales. Banskota et al. (2011) found energy features performed poorly compared to both spectral bands and wavelet coefficients for pine species discrimination. Similar to species discrimination, narrow band reflectance at some specific wavelength regions (such as red-edge and NIR water absorption regions) have been found to be greatly sensitive to variation in LAI (Asner, 1998). Hence, the selection of the appropriate coefficients offers better utilization of the

strength of the wavelength transform and might provide improved estimates of LAI. In addition, objective variable selection also facilitates direct comparison between the original spectral bands and wavelet coefficients with respect to their significance to LAI estimation.

The major objective of this study was to assess the utility of Haar wavelet coefficients for estimating LAI in different vegetation types: deciduous, mixed and coniferous. A genetic algorithm was used to identify the wavelet coefficients that provided the best LAI estimates.

2. Background on wavelet transforms

A detailed introduction to the Haar discrete wavelet transform (DWT) and its scientific applications can be found in Walker (1999). Here, we limit ourselves to only the basic concepts of the Haar DWT considered relevant to this study.

A wavelet transform enables signal (data) analysis at different scales or resolutions by creating a series of shifted and scaled versions of the mother wavelet function (Banskota et al., 2011; Hsu, 2007). The term “mother” implies that a set of basis functions $\{\psi_{a,b}(\lambda)\}$ can be generated from one main function, or the mother wavelet $\psi(\lambda)$ by the following equation (Bruce et al., 2001),

$$\psi_{a,b}(\lambda) = \frac{1}{\sqrt{a}}\psi\left(\frac{\lambda - b}{a}\right) \quad (1)$$

where *a* is the scaling factor of a particular basis function and *b* is the translation variable along the function’s range. Like all wavelet transforms, the Haar wavelet

transform separates a discrete signal **f** of length *n* into two subsignals, each with length *n/2*. Walker (1998) refers to one of these as a running average and the other as a running difference. We refer to the running average, or trend, as the approximation vector, **a**. For each *m* = 1, 2, 3,…, *n*/2, the approximation coefficient is calculated as shown in Eq. 2.

$$a_m = \frac{f_{2m-1}+f_{2m}}{\sqrt{2}} \tag{2}$$

We refer to the running difference, which Walker (1998) also calls a fluctuation, as the detail vector, **d**. Each of the detail coefficients is calculated as shown in Eq. 3 (*m* similarly defined).

$$d_m = \frac{f_{2m-1}-f_{2m}}{\sqrt{2}} \tag{3}$$

The subsignals of the original signal define the first level of the Haar transform, usually referred to as 1-level. As such, the approximation coefficients and detail coefficients from the first level can be referred to as $\mathbf{a}_1$ and $\mathbf{d}_1$, respectively. Computation of approximation and detail coefficients for subsequent levels can be computed recursively by applying Eqs. 2 and 3 to the previously level's approximation coefficients. Since the subsignals have half the length of the previous signal, $\mathbf{a}_2$ and $\mathbf{d}_2$ will have half the length of $\mathbf{a}_1$, or *n*/4. The number of times *n* is divisible by two yields the possible number of decomposition levels. For example, if $n = 16 = 2^4$, four levels of Haar transforms can be computed.

An efficient way to implement the Haar DWT scheme using filters was developed by Mallat (1989). The basic idea behind this fast algorithm is to represent the wavelet basis as a set of high-pass and low-pass filters. The high-pass and low-pass filters are related (their power sum is equal to one) and called quadrature mirror filters. Thus, the 1-level DWT decomposition of a signal splits it into a low pass version (approximation coefficients) and a high pass version (detail coefficients). The 2-level decomposition is performed on the low pass signal obtained from the first level of decomposition. The final results of a DWT decomposition of a spectrum are sets of wavelet coefficients, with each wavelet coefficient directly related to the amount of energy in the signal at different positions and scales.

3. A brief introduction to genetic algorithms

Holland (1975) introduced the genetic algorithm (GA) as a method for solving optimization problems that mimics ideas from Darwin's biological theory of “natural selection” and “survival of the fittest” (Lin and Sarabandi, 1999). In nature, the fitness of individuals depends on their genes and those individuals with greater fitness have a greater chance of survival. In a genetic algorithm for variable selection, the fitness of random subsets of variables is evaluated. Subsets with greater fitness are allowed to survive and undergo exchange of variables. Genetic algorithm has been found useful for selecting variables for different remote sensing applications (Fang et al., 2003; Kooistra et al., 2003; Luo et al., 2003; Siedlecki and Sklansky, 1989; Ulfarsson et al., 2003; Vaiphasa et al., 2007).

A genetic algorithm is initialized with input parameters and a random population of a subset of variables. Each subset of variables is called the ‘individual’ and the pool of all

tested individuals is the ‘population’. Each individual is assessed according to a specified fitness function and the fitness value is calculated. The individuals with fitness lesser than the average are discarded. At this point, the population is shrunk to half its original size and genetic algorithm cross-breeds the retained individuals to replace the discarded individuals. All the individuals are given a chance for random mutation. The population returns to the original size and the process can continue again at the fitness evaluation step. The entire process stops once a predefined criterion is met or convergence is reached and best subset of identified variables is returned.

4. Methods

4.1 Description of the study region

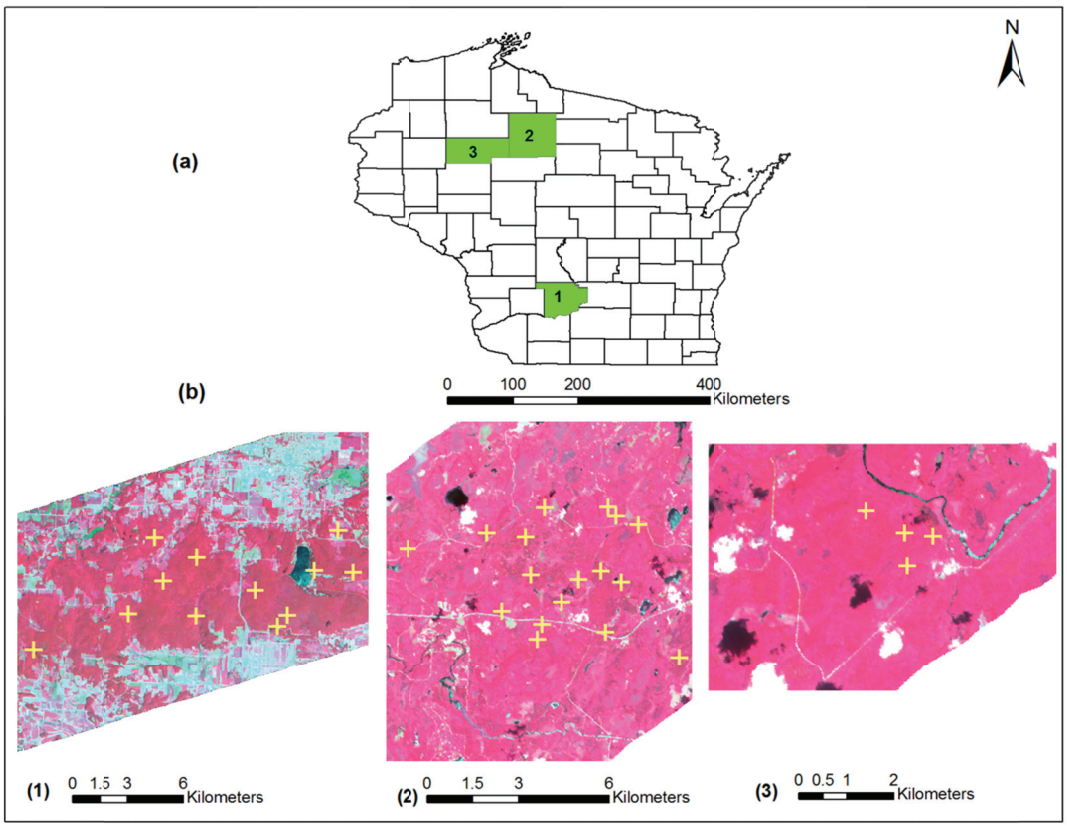

Figure 1: Study area: (a) State of Wisconsin, USA, showing general location of the three field sites and (b) plot locations overlaid on a false color composite from the AVIRIS image (red: 860 nm; green: 667 nm and blue: 540 nm)

The study area comprises a range of coniferous, deciduous and mixed forest types across several ecoregions within the state of Wisconsin, USA (Figure 1). The northern-most forest sites were located within the Northern Lakes and Forest ecoregion (90.120 °W, 45.936 °N to 90.318 °W, 45.966 °N), which is dominated by a mixed-hardwood forest originating from large-scale clear-cut practices of the early 20th century (Curtis, 1959). The overstory vegetation was comprises mostly of northern hardwoods dominated by sugar maple (*Acer saccharum*), basswood (*Tilia americana*), and white ash (*Faxinus americana*). The dominant coniferous species are balsam fir (*Abies balsamea*), white pine (*Pinus strobus*), red pine (*Pinus resinosa*), Tamarack (*Larix laricina*), and black spruce (*Picea mariana*).The remaining northern forest sites were located in the highly fragmented landscape found within the Chequamegon-Nicolet National Forest, near Park Falls, Wisconsin (90.775 °W, 45.624 °N to 90.794 °W, 45.628 °N). Much of this study area is actively managed for multiple-use goals, including wood and fiber, fish and wildlife habitat, and recreation. The dominant vegetation in the upland areas comprises of northern hardwood species including sugar maple (*Acer saccharum*), basswood (*Tilia americana*), green ash (*Fraxinus pennsylvanica*), trembling aspen (*Populus tremuloides*), and paper birch (*Betula papyrifera*). The more poorly drained, lowland deciduous sites are characterized by wetland species such as speckled alder (*Alnus incanta*) as well as hardwood species including black ash (*Fraxinus nigra*) and red maple (*Acer rubra*). The southern sites were located in the Baraboo Hills (89.850 °W, 43.385 °N to 89.640 °W, 43.391°N) which are in the Driftless ecoregions of Wisconsin. Most of the forests in the Baraboo Hills were cleared by the 1870s and have since recovered to forests dominated by red oak

(*Quercus rubra*), white oak (*Quercus alba),* green ash (*Fraxinus pennsylvanica*), white ash (*Fraxinus americana*), bitternut hickory (*Carya cordiformis*), shagbark hickory (*Carya ovata*), hemlock (*Tsuga canadensis*), black spruce (*Picea mariana*), with sugar maple (*Acer saccharum*), red maple *(Acer rubrum*), and basswood (*Tilia americana)* in the uplands.

4.2 LAI measurement protocol

Optical measurements of effective LAI (L_e) were made at a height of one meter above the forest floor using digital hemispherical photography (DHP) (Chen et al., 2006; Jonckheere et al., 2004; Leblanc et al., 2005; Zhang et al., 2005). L_e represents the equivalent leaf area of a canopy with a random foliage distribution to produce the same light interception as the true LAI (Fernandes et al., 2004). The DHP method measures the canopy gap fraction across selected zenith angles from beneath the canopy through a hemispherical (fisheye) lens, oriented towards the zenith. For this study we utilized a Nikon CoolPix 5000 digital camera, leveled on a tripod with an attached Nikon FC-E8 fisheye lens, which has a field of view of 183°. The overstory LAI of the forest stands can be calculated using Eq. 4 (Leblanc and Chen, 2001):

$$\mathrm{LAI} = \frac{Le}{\Omega} - \alpha \qquad (4)$$

where L_e is the effective LAI calculated from DHP, Ω is the clumping index and α is the woody-to-total leaf area ratio ($\alpha = W/Le\,(1/\Omega)$), where W represents the woody-surface-area-index (half the woody area m^{-2} ground area). In this study, we calculated LAI from L_e by only correcting the effect of clumping and neglecting the effect of woods and branches in Eq. 4 (i.e., LAI = L_e/ Ω).

DHP images were collected at 33 plots (60 m X 60 m) characterized by deciduous (18 plots), coniferous (11 plots), and mixed (4 plots) vegetation types, spanning across the study area. All DHP measurements used in this study occurred during the peak of the summer growing seasons in 2008 and 2010 during uniformly overcast skies or during dusk or dawn when the sun was hidden by the horizon. All DHP images were collected in the JPEG format at the highest resolution (2560 x 1920 pixels) to maximize the detection of small canopy gaps (Leblanc et al., 2005). In addition, we collected the images using a technique that maximizes the differentiation between leaf and sky and minimize blooming around leaves (Zhang et al., 2005). For each plot, we measured L_e at nine subplot locations: the plot center, 30 meters from the plot center in each of the four cardinal directions, and the mid-point of each 30m transect. All DHP images were processed using the DHP software (Leblanc et al., 2005), using a nine ring configuration but selecting only the first six rings for analysis to minimize the impacts of large zenith angles on the L_e retrievals and the calculation of LAI (Chen et al., 2006; Leblanc et al., 2005). We calculated an average Ω value for each plot using the combined functionality in the DHP-TRAC software (Leblanc et al., 2005).

4.2 AVIRIS image processing

Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data used in this study (Flight ID: f080713t01 and f080714t01) were acquired in 2008 on NASA's ER-2 aircraft during the month of July (i.e., peak growing season) at an altitude of 20 km, yielding a pixel (i.e., spatial) resolution of approximately 17 m (16.8 - 17.0 m). The AVIRIS instrument produces 224 spectral bands (or wavelengths), with an approximate full-

width half-maximum of 10 nm for each wavelength over the spectral range of 370 to 2500 nm (Green et al., 1998).

All AVIRIS image preprocessing involved four distinct steps: the development of an integrated cloud and cloud-shadow mask, cross-track illumination correction, removal of overlapping bands (between detectors 1 and 2 as well as between 2 and 3), and correction of atmospheric effects and conversion to top-of-canopy (TOC) reflectance. AVIRIS images often have systematic cross-track illumination effects due to a combination of flight path orientation and relative solar azimuth. We removed this effect by developing band-wise bilinear trend surfaces, ignoring all cloud/shadow-masked pixels, and trend-normalizing the images by subtracting the illumination trend surface and adding the image mean to arrive at cross-track illumination corrected bands. Atmospheric correction and conversion of the cross-track illumination corrected images to TOC reflectance was done using the ACORN5b^{TM} software (Atmospheric CORrection Now; Imspec LLC, USA). Due to the low ratio of signal to noise at both spectral ends (366 nm - 395 nm and 2467 nm - 2497 nm), and in bands around the major water absorption regions (1363 nm – 1403 nm and 1811 nm -1968 nm) those wavelength regions were dropped, resulting in a final total of 184 bands. The point map of field sampling plots was overlaid on the AVIRIS image data. The pixel reflectances corresponding to the center of the plot locations were extracted for the final 184 AVRIS bands.

Out of a total 18 deciduous among 33 combined plots, one with the highest LAI (6.67) had unusually low reflectance throughout the near-infrared (NIR) region (maximum reflectance of 42% at NIR plateau). We could not put our finger on what

caused such low NIR reflectance in the plot. It might be due to possible change in vegetation conditions between the time of field measurements and image acquisition or due to error in GPS coordinates. We ran a test based upon Cook's Distance which identifies influential observation on the basis of how entire linear regression function changes when a certain observation is deleted (Cook, 1979). The test identified the plot in question as suspicious (partial F-statistic = 0.83, Cook's distance = 0.969) by its very high leverage value on regression parameter estimates. The coefficient of determination (R^2) improved to 0.72 from 0.32 after removing that plot. This outlier was thus removed from further analysis.

4.3 Calculation of discrete wavelet coefficients

The DWT was implemented in Matlab (version 7.4, The Mathworks, Inc.) using a dyadic filter tree as described in Section 2. The hyperspectral signal in the spectral domain extracted for each pixel location was passed through a series of low pass and high pass filters related to Haar wavelets. The decomposition level was chosen such that it was maximized (6 for 184 bands using the Haar wavelet). All the detail wavelet coefficients calculated at each level and approximation coefficients at final level were concatenated to produce a final wavelet dataset.

4.4 Variable selection by GA

The genetic algorithm was set up using the GA toolbox in Matlab (version 7.1, The Mathworks, Inc.). The algorithm was run separately for two datasets (combined plots and deciduous plots) to find the optimum set of features. The fitness function used was the leave-one-out cross validation (CV-RMSE) between observed LAI and estimated LAI. The goal of the algorithm was to find the subsets of variables that minimize the

fitness function. We used GA to find the optimal subset containing different numbers of features (two to six). To avoid multicollinearity, the CV-RMSE was set to one for subsets with highly correlated variables (correlation coefficient greater than 0.8), ensuring that their fitness was minimized.

The algorithm was initially run for multiple times, and the best parameter set-up suited to this study was determined as following: (1) population = 100, (2) mutation rate = 0.5, (3) cross over rate = 0.5, and (4) stopping criterion = 100 generations or 25 generations with no improvement in the best fitness value. With these parameters, GA was run five times for each dataset to find the best subsets. The subsets that provided the minimum fitness function were chosen for further statistical analysis.

4.5 Statistical analysis

Multiple linear regressions were used to create regression models for LAI in deciduous and combined plots as a function of the wavelet variables and spectral bands. Two subsets of variables were used as independent variables for the combined sets of plots: (1) wavelet coefficients selected by GA (WAVE_ALL) and (2) spectral bands selected by GA (BANDS_ALL). Similarly, two subsets of variables were considered for deciduous plots: (1) wavelet coefficients (WAVE_DECI) and (2) spectral bands selected by GA (BANDS_DECI). Selected models were subject to the following constraints: (1) all independent variables were significant at $\alpha = 0.05$ for all plots and at a relaxed $\alpha = 0.15$ for deciduous plots (because of only 17 observations), and (2) there could not be multi-collinearity, i.e., all variance inflation factors (VIFs) were less than 10. Consideration was also given to model parsimony, i.e., a model with fewer variables was preferred to one with many variables. Once the best models were chosen, leave-

one-out cross validation was used to validate them. In each iteration of the cross validation, one of the observations was set aside, a predictive equation was calculated using the remaining observations, and the single observation set aside was used to validate the model. This procedure was repeated for all observations. The cross-validation coefficient of determination (CV-R^2) and cross-validation RMSE (CV-RMSE) were calculated to assess the prediction capabilities of the best models.

5. Results

5.1 Wavelet coefficients and spectral bands selection by GA

Genetic algorithm was initially used to select five best subsets with two to six variables for each dataset. The results for wavelet coefficients and spectral bands for combined plots are shown in Table 1 and 2 respectively. It can be seen in Table 1 that the model with five coefficients provided the best accuracy (adjusted R^2 = 0.75, CV-R^2 = 0.71, CV-RMSE = 0.46). The variables were significant at $\alpha = 0.05$ and VIF of all variables were below five. Based upon the results, we chose the wavelet subset with five variables as the final subset for WAVE_ALL.

Table 2 shows that the five bands provided the best accuracy (adjusted R^2 = 0.60, CV-R^2 = 0.52, CV-RMSE = 0.59). However, there was no significant difference in both CV- RMSE and CV-R^2 between the models with five and two variables. Significance testing showed that none of the models with greater than two variables were significant ($\alpha = 0.15$). Hence, we chose the two spectral bands for the combined set of plots as the best subset (BANDS_ALL). With similar analysis, we chose four wavelet coefficients for deciduous plots (WAVE_DECI), and four spectral bands for deciduous plots (BANDS_DECI) as final subsets.

Table 1: Model statistics for best models with different number of wavelet coefficients (2 to 6) selected by genetic algorithm. CV-RMSE and CV-R^2 refer to the leave-one-out cross validation RMSE and R^2 respectively.

Statistics	2 variable	3 variable	4 variable	5 variable	6 variable
RMSE	0.63	0.57	0.50	0.43	0.41
CV-RMSE	0.66	0.61	0.57	0.46	0.46
R^2	0.50	0.60	0.70	0.79	0.81
CV-R^2	0.40	0.49	0.56	0.71	0.70
Adjusted R^2	0.47	0.56	0.64	0.75	0.77

Table 2: Model statistics for best models with different number of spectral bands (2 to 6) selected by genetic algorithm. CV-RMSE and CV-R^2 refer to the leave-one-out cross validation RMSE and R^2 respectively.

Statistics	2 variable	3 variable	4 variable	5 variable	6 variable
RMSE	0.57	0.57	0.55	0.55	0.55
CV-RMSE	0.60	0.60	0.59	0.59	0.62
R^2	0.58	0.60	0.64	0.66	0.67
CV-R^2	0.50	0.50	0.52	0.52	0.50
Adjusted R^2	0.55	0.56	0.59	0.60	0.60

The five wavelet coefficients selected by WAVE_ALL subset were fine scale detail coefficients corresponding to first and second levels of decomposition (two from 1-level, three from 2-level). Figure 2 shows the plots of 1-level and 2-level DWT coefficients for two deciduous plots (LAI = 2.98 and LAI = 5.66) and the location of the selected wavelet coefficients relative to the original spectral bands. The two coefficients selected from 1-level were related to 1120 nm - 1130 nm and 2208 nm - 2228 nm wavelength regions. The three coefficients from 2-level were related to 714 nm - 743 nm, 1110 nm - 1139 nm, and 2198 nm - 2238 nm wavelength regions.

The four wavelet coefficients selected by WAVE_DECI subset were three fine-scale detail coefficients from 1-level and 2-level and one coarse scale detail coefficient from 5-level. Two coefficients from 1-level were related to 1120 nm - 1130 nm and 1273 nm - 1293 nm, and one coefficient from 1-level was related to 1263 nm - 1303 nm wavelength regions respectively. The coarse scale coefficient at 5-level corresponded to the broader wavelength region spanning from 424 nm to 724 nm.

Two spectral bands of wavelength 841 nm and 2437 nm were selected by BANDS_ALL and four spectral bands of wavelength 424 nm, 995 nm, 1353 nm, and 2467 nm were selected by BAND_DECI.

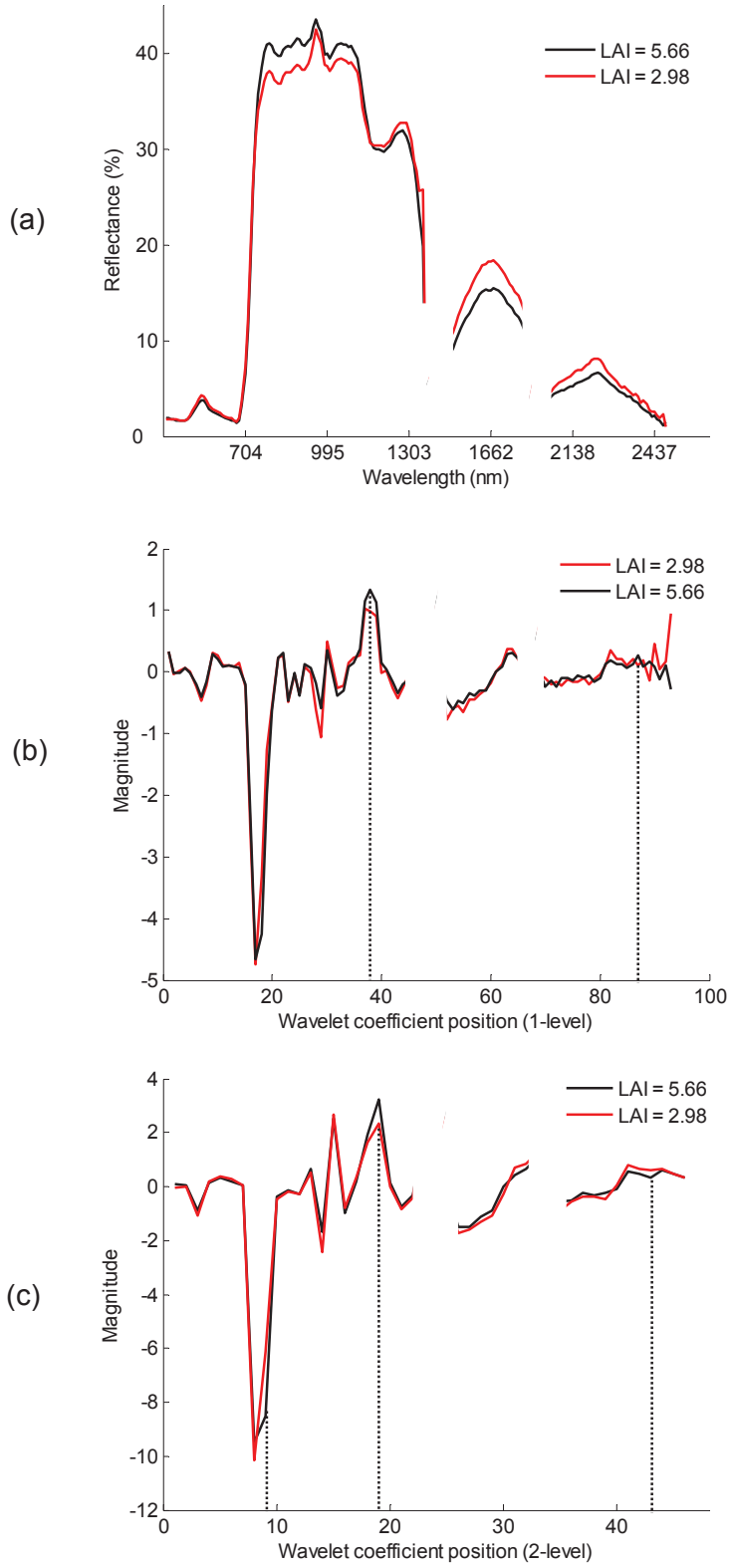

Figure 2: (a) AVRIS reflectances from two plots in deciduous stands with LAI = 2.98 (red) and LAI = 5.66. (b) Detail Haar wavelet coefficients spectra at 1-level coinciding with panel (a). Note that exact coincidence is not possible because there are 92 coefficients and 184 spectral bands. (c) Detail Haar wavelet coefficients at 2-level coinciding with panel (a) and (b). The locations of the coefficients selected for combined plots are shown with dashed lines in panel B and C.

5.2 Regression results

The final results of linear regression analyses are presented in Table 3. The first column shows the type of independent variables used in the model. The second and third columns show the number of plot observations and number of variables used in the model. Fourth to last columns show the model coefficient of determination (R^2), adjusted R^2, cross-validation R^2 (CV-R^2), RMSE and cross validation RMSE (CV-RMSE) respectively. The results show that the model derived from WAVE_ALL provided the best fit when the regression models were built using observations from all plots (CV-R^2 = 0.71, CV-RMSE = 0.46). For deciduous plots only, WAVE_DECI provided better estimates of LAI (CV-R^2 = 0.79, CV-RMSE = 0.31) than BANDS_DECI (CV-R^2 = 0.69, CV-RMSE = 0.44). The accuracy of LAI estimates from BANDS_DECI and WAVE_ALL were similar. The relationships between observed and best predicted LAI for all subsets are shown in Figures 3-6.

Table 3: Regression results between observed LAI and estimated LAI from four different subsets of data. CV- R^2 and CV-RMSE refers to the leave-one-out cross validation R^2 and RMSE respectively.

Features	Plots	Variables (n)	R^2	Adjusted R^2	CV- R^2	RMSE	CV-RMSE
WAVE_ALL	32	5	0.79	0.75	0.71	0.43	0.46
BANDS_ALL	32	2	0.58	0.52	0.50	0.58	0.60
BANDS_DECI	17	4	0.80	0.73	0.69	0.36	0.44
WAVE_DECI	17	4	0.90	0.86	0.79	0.28	0.31

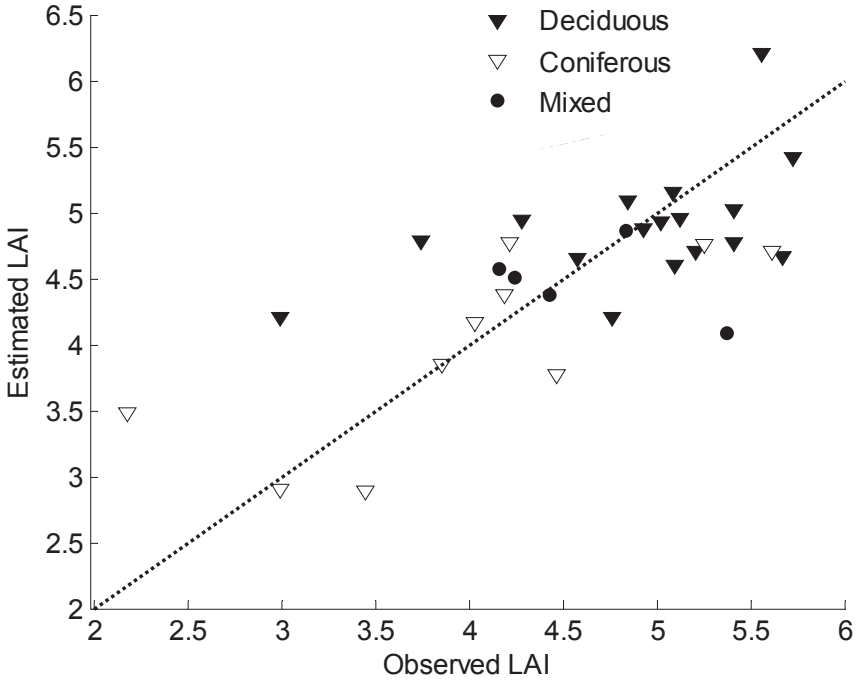

Figure 3: Observed versus estimated LAI from spectral bands for combined plots (BANDS_ALL).

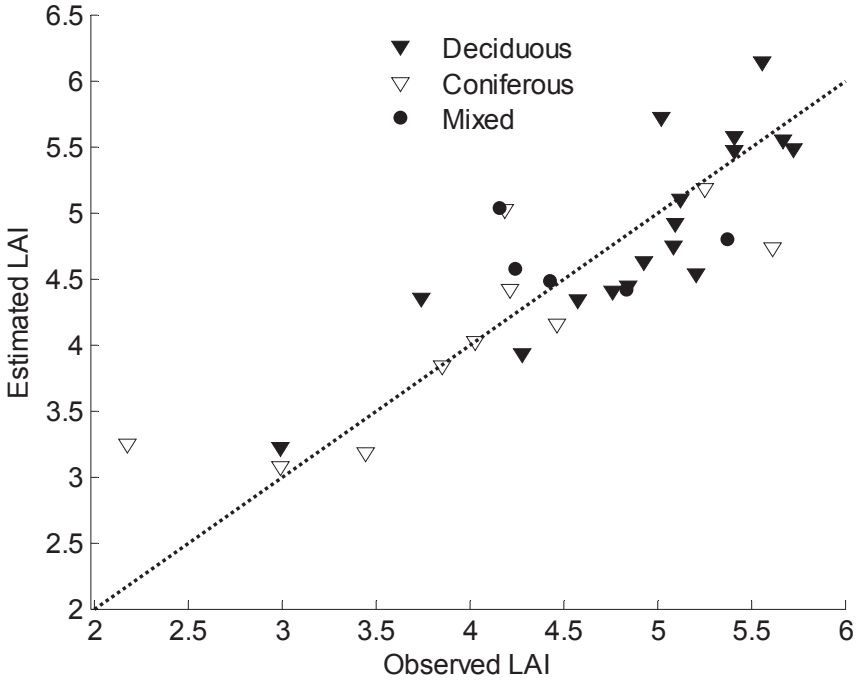

Figure 4: Observed versus estimated LAI from wavelet coefficients for combined plots (WAVE_ALL).

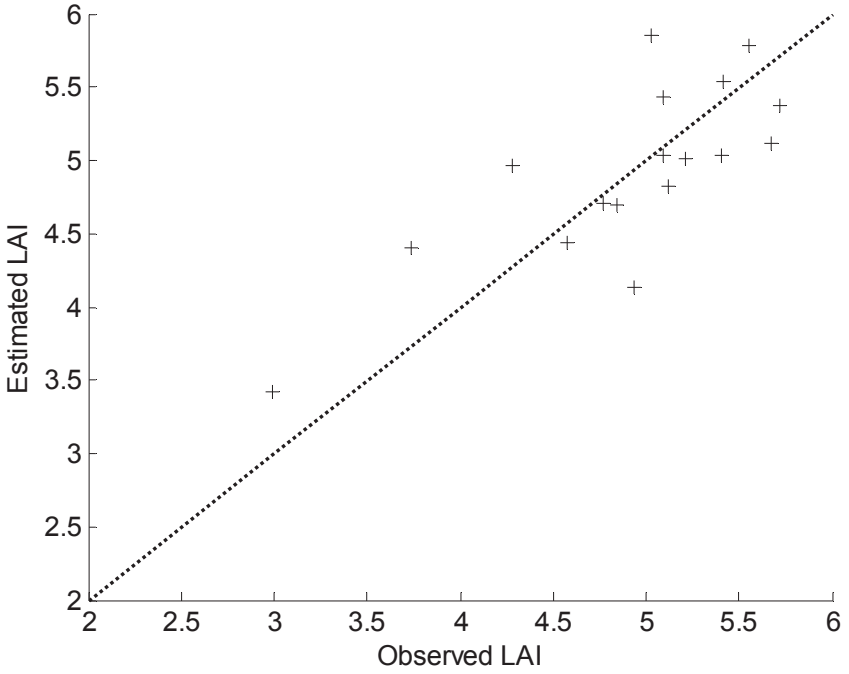

Figure 5: Observed versus estimated LAI from spectral bands for deciduous plots (BANDS_DECI).

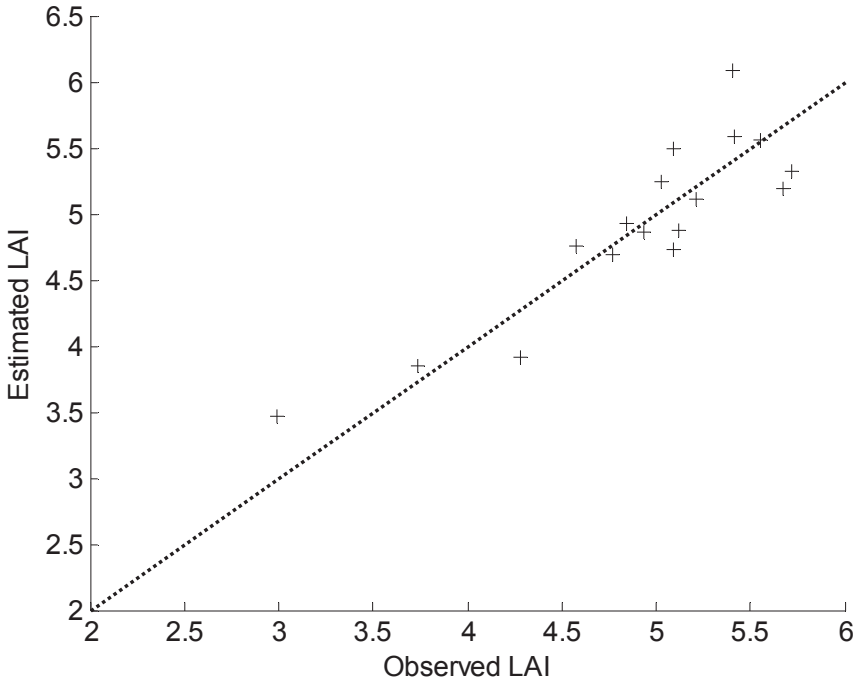

Figure 6: Observed versus estimated LAI from wavelet coefficients for deciduous plots (WAVE_DECI).

6. Discussion

Previous studies have used different combinations of bands in spectral vegetation indices (SVI) for improved LAI estimation (Broge and Leblanc, 2000; Brown et al., 2000; Eklundh et al., 2001).Other studies have reported better performance by models with multiple bands than by univariate models using SVIs (Darvishzadeh et al., 2008; Lee et al., 2004). Gong et al. (1995) and Johnson et al. (1994) found the first and second derivatives of the reflectance less sensitive to background effects and more useful for LAI prediction than spectral bands. Haar DWT coefficients in this study might have potentially combined the strengths of SVI, derivatives, and spectral bands for LAI estimation. First level detail coefficients from the Haar DWT are functionally equivalent to first derivatives of the reflectance data (Bruce et al., 2002). On the other hand, higher level detail coefficients are similar to some SVIs as they tend to measure the contrast over a broad spectral interval (for e.g., between green and red bands, red band and red-edge region, etc.). Additionally, these coefficients exhibit less correlation and reduced noise compared to spectral bands and hence are more suitable for linear regressions. The regression models with wavelet subsets might have benefitted from these advantages accrued from wavelet transform and provided better estimates of LAI than spectral bands.

Canopy reflectance around the red-edge (704 nm - 724 nm) and within 1275 nm - 1350 nm in the NIR part of the reflectance spectrum has been found to be greatly sensitive to the changes in LAI (Asner, 1998; Jacquemoud et al., 2009; Lee et al., 2004). In this study, WAVE_DECI selected fine scale coefficients near both wavelength regions. It also selected one coarse scale coefficient (5-level) related to bands in the

visible through red-edge region. The coarse scale coefficient was equivalent to the difference between 4-level approximation coefficients from the first 16 (424 nm - 560 nm) and the next 16 bands (570 nm - 724 nm) . There is no reported implication of the spectral differences between these broad band regions in the literature. But the results show that such coarse scale differences in combination with fine scale differences as detected by other coefficients in red-edge and NIR region might be important for explaining LAI variation in deciduous vegetation types.

Previous studies reported poor accuracy of LAI estimates in areas with mixed vegetation and recommended vegetation-type-specific models to estimate LAI from remote sensing data (Colombo et al., 2003; Fassnacht et al., 1997; Turner et al., 1999). One of the major goals of remote sensing is to build empirical models applicable over a range of vegetation condition and vegetation types. Such models potentially require less-intensive ground sampling efforts and significantly reduce both time and cost. The results in this study showed promise for building such a unified regression model between LAI and AVRIS data with the aid of Haar DWT. The best model with spectral bands for combined plots (BANDS_ALL) used two bands in the model and estimated LAI with poor accuracy (CV-R^2 = 0.50; CV-RMSE = 0.60). The accuracy of LAI estimates was greatly improved by the model using WAVE_ALL (CV-R^2 = 0.71; CV-RMSE = 0.46). The accuracy was comparable to the accuracy of the best model for deciduous stands using spectral bands (BANDS_DECI, CV-R^2 = 0.69; CV-RMSE = 0.44). It indicates that similar levels of prediction accuracy for LAI in combined vegetation types can be achieved by using wavelet coefficients to that achieved by spectral bands in a deciduous vegetation type. WAVE_ALL selected only fine scale

coefficients from 1-level and 2-level decompositions. Three of these coefficients were similar to the fine scale coefficients selected by WAVE_DECI. In contrast to the latter, WAVE_ALL did not select any coarse scale coefficients, but additionally selected two coefficients related to 2198 nm - 2238 nm in the SWIR part of the spectrum. The strong relationship between SWIR bands and LAI has previously been suggested by many studies (Brown et al., 2000; Darvishzadeh et al., 2008; Eklundh et al., 2001). The selected fine scale detail coefficients in the SWIR, NIR and red-edge region are related to the reflectance difference over neighboring bands and might be less sensitive to the reflectance amplitude in those bands. It indicates that WAVE_ALL selected the fine scale coefficients that might be similarly sensitive to change in LAI in different functional types (deciduous – coniferous) and thus suppressed the background reflectance variation due to difference in vegetation types.

Pu and Gong (2004) found energy features from wavelet transforms of hyperspectral data useful for LAI estimation compared to spectral bands selected by stepwise regression. Though the energy features drastically reduce the number of dimensions, they average all the detail coefficients at a particular scale into a single feature. For example, all the 1-level detail coefficients are squared and summed up into one energy feature. As we can see from Figure 2 (b) and 2 (c), not all the 1-level and 2-level coefficients are sensitive to change in LAI from LAI = 2.98 to LAI = 5.96. Hence, the selection of useful coefficients related to different scales and positions that explains maximum variation in LAI is necessary for building an accurate predictive model using wavelet transformed hyperspectral data.

7. Conclusion

The main objective of this study was to investigate the utility of the Haar DWT in estimating LAI in mixed vegetation types using AVIRIS hyperspectral data. DWT transforms the hyperspectral data into wavelet features at a variety of spectral scales. The multiscale features detect and isolate variation in the reflectance continuum not detectable in the original reflectance domain such as amplitude variations over broad and narrow spectral regions. This study demonstrated that such wavelet features at different scales show increased sensitivity to variations in LAI and can provide improved LAI estimates compared to the original reflectance data. While the regression model for deciduous types utilized both coarse scale and fine scale features related to visible-red edge and NIR reflectance, the model for combined vegetation types used the fine scale features only related to red-edge, NIR and SWIR reflectance. Coarse scale features are potentially insensitive to background variation caused by different vegetation types and thus are more useful for explaining variation caused by LAI.

This study highlighted the importance of the wavelet features for empirical estimation of LAI. Additional information could be obtained by running a sensitivity analysis to analyze the effect of variation in LAI on wavelet features at different positions and scales. Such a study might also help identify features that are more sensitive to LAI and less to background signals caused by soil, crown cover, etc.

The theory of wavelet transforms is evolving, and more wavelet families are being introduced. Different families of wavelets are suited to different signals and applications. Hence a comparative analysis of Haar with other families of wavelets might improve LAI estimation using AVIRIS data even further.

References

Asner, G.P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, 64, 234-253.

Banskota, A., Wynne, R.H., & Kayastha, N. (2011). Improving within-genus tree species discrimination using the discrete wavelet transform applied to airborne hyperspectral data. *International Journal of Remote Sensing*, 32, 3551-3563.

Blackburn, G.A. (2007). Wavelet decomposition of hyperspectral data: A novel approach to quantifying pigment concentrations in vegetation. *International Journal of Remote Sensing,* 28*,* 2831–2855.

Blackburn, G.A., & Ferwerda, J.G. (2008). Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of Environment,* 112, 1614-1632.

Bradshaw, G.A., & Spies, T.A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *Journal of Ecology,* 80, 205-215.

Broge, N.H., & Leblanc, E. (2000). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment,* 76, 156–172.

Brown, L., Chen, J.M., Leblanc, S.G., & Cihlar, J. (2000). A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: An image and model analysis. *Remote Sensing of Environment*, 71, 16–25.

Bruce, L.M., Morgan, C., & Larsen, S. (2001). Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms. *IEEE Transactions on Geoscience and Remote Sensing,* 39, 2217-2226.

Bruce, L.M., Koger, C.H. & Jiang, L. (2002). Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 2331–2338.

Chen, J.M., Pavlic, G., Brown, L., Cihlar, J., Leblanc, S.G., White, H.P., Hall, R.J., Peddle, D.R., King, D.J., & Trofymow, J.A. (2002). Derivation and validation of Canada-wide coarse-resolution leaf area index maps using high-resolution satellite imagery and ground measurements. *Remote Sensing of Environment,* 80, 165–184.

Chen, J.M., Govind, A., Sonnentag, O., Zhang, Y.Q., Barr, A., & Amiro, B. (2006). Leaf area index measurements at Fluxnet-Canada forest sites. *Agricultural and Forest Meteorology, 140*, 257-268.

Cheng, T., Rivard, B., & Sanchez-Azofeifa, A. (2011).Spectroscopic determination of leaf water content using continuous wavelet analysis. *Remote Sensing of Environment, 115*, 659-670.

Cohen, W.B., Maiersperger, T.K., Gower, S.T., & Turner, D.P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, 84, 561–571.

Colombo, R., Bellingeri, D., Fasolini, D., & Marino, C.M. (2003). Retrieval of leaf area index in different vegetation types using high resolution satellite data. *Remote Sensing of Environment,* 86, 120–131.

Cook, R.D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, 74, 169-174.

Coops, N.C., Smith, M.L., Martin, M.E., & Ollinger, S.V. (2003). Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions*, 41, 1338- 1346.

Curran, P.J. (1994). Imaging spectrometry. *Progress in Physical Geography*, 18, 247-266.

Curtis, J.T. (1959). *The vegetation of Wisconsin: An ordination of plant communities*. (2nd ed.). Madison: The University of Wisconsin Press.

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in heterogeneous grassland. *Remote Sensing of Environment*, 112, 2592-2604.

Eklundh, L., Harrie, L., & Kuusk, A. (2001). Investigating relationships between Landsat ETM+ sensor data and leaf area index in a boreal conifer forest. *Remote Sensing of Environment, 78*, 239-251.

Fang, H., Liang, S., & Kuusk, A. (2003). Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. *Remote Sensing of Environment,* 85, 257–270.

Fassnacht, K. S., Gower, S. T., MacKenzie, M. D., Nordheim, E. V., & Lillesand, T. M. (1997). Estimating the leaf area index of north central Wisconsin forest using the Landsat Thematic Mapper. *Remote Sensing of Environment*, 61, 229–245.

Fernandes, R., Miller, J.R., Chen, J.M., & Rubinstein, I.G. (2004). Evaluating image-based estimates of leaf area index in boreal conifer stands over a range of scales using high-resolution CASI imagery. *Remote Sensing of Environment*, 89, 200–216.

Geladi, P., & Kowalski, B.R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta, 185*, 1-17.

Gong, P., Pu, R.L., & Miller, J.R. (1995). Coniferous forest leaf-area index estimation along the Oregon transect using compact airborne spectrographic imager data. *Photogrammetric Engineering & Remote Sensing*, 61, 1107-1117.

Goward, S.N., Huemmrich, K.F., & Waring, R.H. (1994). Visible-near infrared spectral reflectance of landscape components in western Oregon. *Remote Sensing of Environment, 47*, 190-203.

Gower, S.T., Kucharik, C.J., & Norman, J.M. (1999). Direct and indirect estimation of leaf area index, F(Apar), and net rimary production of terrestrial ecosystems. *Remote Sensing of Environment, 70*, 29-51

Gower, S. T., Krankina, O., Olson, R. J., Apps, M., Linder, S., & Wang, C. (2001). Net primary production and carbon allocation patterns of boreal forest ecosystems. *Ecological Applications*, 11, 1395−1411.

Gower, S. T. (2003). Patterns and mechanisms of the forest carbon cycle. *Annual Review of Environment and Resources*, 28, 169−204.

Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., & Chippendale, B.J. (1998). Imaging spectroscopy and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, 65, 227–248.

Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., & Strachan, I.B. (2004). Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment, 90*, 337-352

Hansen, P.M., & Schjoerring, J.K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86, 542–553.

Holland, J.H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan.

Hsu, P.H. (2007). Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry & Remote Sensing*, 62, 78-92.

Huemmrich, K.F., Privette, J.L., Mukelabai, M., Myneni, R.B., & Knyazikhin, Y. (2005). Time-series validation of MODIS land biophysical products in a Kalahari woodland, Africa. *International Journal of Remote Sensing,* 26, 4381-4398.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., & Ustin, S.L. (2009). PROSPECT+ SAIL models: a review of use for vegetation characterization, *Remote Sensing of Environment,* 113, S56–S66.

Johnson, L.F., Hlavka, C.A., & Peterson, D.L. (1994). Multivariate analysis of AVIRIS data for canopy biochemical estimation along the Oregon transect. *Remote Sensing of Environment,* 47, 216-230.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., & Baret, F. (2004). Review of methods for in situ leaf area index determination - Part I. theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, 121, 19-35.

Kaewpijit, S., Le Moigne, J., & El-Ghazawi, T. (2003). Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 863–871.

Koetz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment*, 95,115−124.

Kooistra, L., Wanders, J., Epema, G.F., Leuven, R.S.E.W., Wehrens, R., & Buydens, L.M.C. (2003). The potential of field spectroscopy for the assessment of sediment properties in river floodplains. *Analytica Chimica Acta*, 484, 189–200.

Leblanc, S.G., & Chen, J.M. (2001). A practical scheme for correcting multiple scattering effects on optical LAI measurements. *Agricultural and Forest Meteorology*, 110, 125-139.

Leblanc, S.G., Chen, J.M., Fernandes, R., Deering, D.W., & Conley, A. (2005). Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agricultural and Forest Meteorology*, 129, 187-207.

Lee, K.S., Cohen, W.B., Kennedy, R.E., Maiersperger, T.K., & Gower, S.T. (2004). Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment*, 91, 508–520.

Liu, M.L., Liu, X.N., Ding, W.C., & Wu, L. (2011). Monitoring stress levels on rice with heavy metal pollution from hyperspectral reflectance data using wavelet-fractal analysis. *International Journal of Applied Earth Observation and Geoinformation,* 13, 246-255.

Lin, Y.C., & Sarabandi, K. (1999). Retrieval of forest parameters using a fractal-based coherent scattering model and a genetic algorithm. *IEEE Transactions on Geoscience and Remote Sensing,* 37, 1415-1424.

Luo, J.C., Zheng, J., Leung, Y., & Zhou, C.H. (2003). A knowledge integrated stepwise optimization model for feature mining in remotely sensed images. *International Journal of Remote Sensing*. 24, 4661–4680.

Mallat, S. (1989). A theory for multi-resolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674–693.

Malinowski, Z., Martin, E., Homolová, L., Gastellu-Etchegorry, J.P., Zurita-Milla, R., Schaepman, M.E., Pokorný, R., Clevers, J.G.P.W., & Cudlín, P. (2008). Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution. *Remote Sensing of Environment*, 112, 1-18.

Meroni, M., Colombo, R., & Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment*, 92, 195-206.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., & Running, S. (2002). Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of Environment*, 83, 214−231.

Oren, R., Ewers, B.E., Todd, P., Phillips, N., & Katul, G. (1998). Water balance delineates the soil layer in which moisture affects canopy conductance. *Ecological Applications*, 8, 990-1002.

Pu, R., & Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, 91, 212-224.

Schaepman, M.E., Koetz, B., Schaepman-Strub, G., & Itten, K.I. (2005). Spectrodirectional remote sensing for the improved estimation of biophysical and -chemical variables: two case studies. *International Journal of Applied Earth Observation and Geoinformation*, 6, 271-282.

Schmidt, K.S., & Skidmore, A. K. (2003). Spectral discrimination of vegetation types in a coastal wetland. *Remote Sensing of Environment*, 85, 92-108.

Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, 10, 335–347.

Thenkabail, P. S., Smith, R. B., & De Pauw, E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, 71, 158–182.

Turner, D.P., Cohen, W.B., Kennedy, R.E., Fassnacht, K.S., & Briggs, J.M. (1999). Relationships between leaf area Index and Landsat TM spectral vegetation Indices across three temperate zone Sites. *Remote Sensing of Environment, 70*, 52-68.

Turner, D.P., Ritts, W.D., Cohen, W.B., Gower, S.T., Zhao, M., Running, S.W., Wofsy, S.C., Urbanski, S., Dunn, A.L., & Munger, J.W. (2003). Scaling Gross Primary Production (GPP) over boreal and deciduous forest landscapes in support of MODIS GPP product validation. *Remote Sensing of Environment*, 88, 256-270.

Ulfarsson, M. O., Benediktsson, J. A, & Sveinsson, J. R. (2003). Data fusion and feature extraction in the wavelet domain. *International Journal of Remote Sensing, 24*, 3933-3945.

Vaiphasa, C., Skidmore, A.K., de Boer, W.F., & Vaiphasa, T. (2007). A hyperspectral band selector for plant species discrimination. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 225-235.

Walker, J.S. (1999). *A primer on wavelets and their scientific applications*. Boca Raton: CRC.

Zhang, Y.Q., Chen, J.M., & Miller, J.R. (2005). Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology*, 133, 166-181.

Zhang, J., Rivard, B., Sanchez-Azofeifa, A., & Casto-Esau, K. (2006). Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. *Remote Sensing of Environment,* 105, 129-141.

CHAPTER 4

AN INNOVATIVE TECHNIQUE FOR BUILDING A LOOK-UP TABLE FOR EFFICIENT INVERSION OF DART TO ESTIMATE FOREST LAI

Abstract

Inversion of a three dimensional (3-D) model such as the Discrete Anisotropy Radiative Transfer (DART) model to estimate canopy variables, has become more computationally efficient with the development of the look-up-table inversion (LUT) technique. A pressing research priority in LUT-based inversion for a complex 3-D model is to determine the optimal LUT grid size defined by the number, ranges, and increments of model inputs. In this study, we present a simple and computationally efficient approach for populating a LUT database with DART simulations over a large number of spectral bands given the availability of (1) equivalent spectra from *in situ* or airborne sensors, and (2) site-specific parameter ranges. In the first step, we initially built a preliminary LUT using model parameters with coarse increments and wide ranges to simulate reflectance for four narrow bands (550 nm, 1139 nm, 1692 nm, and 2208 nm). In the second step, the preliminary LUT was compared with the reflectance measured by AVIRIS data, and the optimal input ranges and realistic parameter combinations that led to simulations close to image spectra were then identified. In the third step, this information was combined with a sensitivity study, and a final LUT was built for the full spectrum of AVIRIS bands in the third step. The final LUT was inverted to estimate leaf area index (LAI) in northern temperate forests using AVRIS data. We compared two different techniques (based upon median and least spectral angle criteria) for choosing a final estimation of LAI from multiple solutions. The results

showed that the inversion solution based upon the median criterion provided better estimates of LAI (RMSE = 0.60) than the solution provided by least spectral angle (RMSE = 0.76). The results indicate that the approach used in this study to find realistic parameter combinations and optimal input increments can be a useful strategy to estimate LAI accurately by DART model inversion.

Keywords**:** *LAI, AVIRIS, DART, look-up-table, inversion*

1. Introduction

Leaf area index (LAI) is a basic vegetation canopy property that controls and moderates different climatic and ecological functions (Huemmrich et al., 2005; Malenovski et al., 2008). Since LAI is one of the principal factors controlling canopy reflectance (Asner, 1998), remote sensing techniques offer practical means of estimating LAI at the landscape level (Gong et al., 1995; Koetz et al., 2005; Schaepman et al., 2005). The majority of studies estimating LAI have employed empirically-based techniques, which are site-specific and may not be extendable to operational use, given different structural and climatic conditions (Curran, 1994; Gobron et al., 1997). On the other hand, the physical approach using canopy reflectance models is based upon an understanding of the physical laws governing the transfer of solar radiation in vegetative canopy; it works with variable measurement conditions, and is better suited for many large-scale applications (Gobron et al., 1997; Kimes et al., 2000; Verstraete et al., 1996). Physically-based methods can also make full use of the high dimensional spectral and multiangular information provided by modern sensors (Chopping et al., 2008; Darvishzadeh et al., 2008).

The majority of physically-based models are one-dimensional (1-D) in the sense that canopies vary only with the height above the ground surface. However, most plant stands contain partial cover and generally exhibit horizontal variability in their structural and optical properties, and are thus imprecisely modeled using 1-D models. The most realistic description of the forest canopy reflected radiation can be provided by 3-D models (Gastellu-Etchegorry et al., 1996; Gastellu-Etchegorry et al., 2003; Kimes et al., 2000; Malenovski et al., 2008; Myneni, 1991; Myneni et al., 1992). The models are referred to as 3-D because the extinction and scattering coefficients that define photon interactions are explicit functions of the spatial coordinates (Myneni et al., 1992).

DART is a widely used 3-D model in several vegetation related scientific works (Duthoit et al., 2008; Gascon et al., 2004; Gastellu-Etchegorry et al., 2004; Malenovski et al., 2008). Though it has a large set of input parameters, the number of parameters can be greatly reduced through simple approximation, making it amenable to inversion to estimate model parameters (Gastellu-Etchegorry et al., 1996). Since its first release (Gastellu-Etchegorry et al., 1996), its accuracy, range of applications, and graphical user interface have been significantly improved (Gascon et al., 2004; Malenovski et al., 2008). DART uses simplifying hypotheses for simulating vegetation landscapes by dividing the scene into rectangular cells characterized by the volume and scattering properties of landscape elements. It can accurately model multiple scattering effects within canopies (Gastellu-Etchegorry et al., 2004). Radiative scattering and propagation are simulated with the exact kernel and discrete ordinate approaches. The model output predicts any specified directional sensor response. It was compared and tested within the radiation transfer model intercomparison (RAMI) experiment (Pinty et al., 2004) and

had an unprecedented level of agreement with other candidate models in simulating heterogeneous canopy spectra (Duthoit, 2008). Figure 1 shows an example of a simple DART computer representation of landscape elements. A detailed description about the DART model can be found in Gastellu-Etchegorry et al. (1996) and Gastellu-Etchegorry et al. (2003).

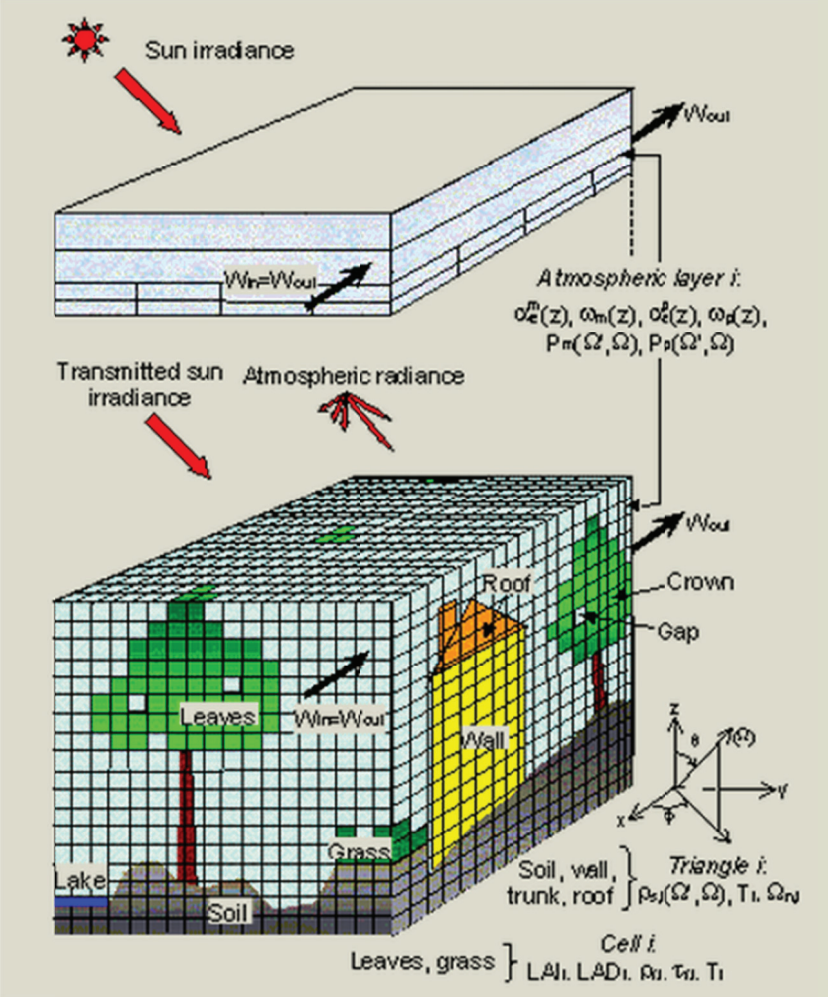

Figure 1: An example of DART computer representation of landscape elements, atmospheric layer, and illumination. The different symbols represent variables that define the optical properties of each DART cell. (Source:http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI3/MODELS/DART/DART.php)

For the estimation of vegetation properties such as LAI, a model has to be inverted against measured reflectance data. Several model inversion techniques are available

(Kimes et al., 1998; Kimes et al., 2000; Schlerf and Atzberger, 2006). Traditional inversion methods like optimization techniques iteratively adjust model parameters until the modeled reflectance 'fits' or matches measured reflectance. These iterative optimization approaches are not feasible for complex models with slow computation time when simulating reflectance over a large number of spectral bands. Recently, the LUT-based inversion methods have been widely used (Darvishzadeh et al., 2008; Hedley et al., 2009; Knyazikhin et al., 1998; Liang, 2007). They require construction of a database of pre-calculated reflectances as a function of key model parameters. Then the spectrum for a particular image pixel is compared with each spectrum in the database, and the closest match to the image spectrum is found. The parameter combination that yields the closest spectrum in the database is considered to be the inversion solution (Mobley et al., 2005). A conceptual diagram for a general LUT inversion is shown in Figure 2.

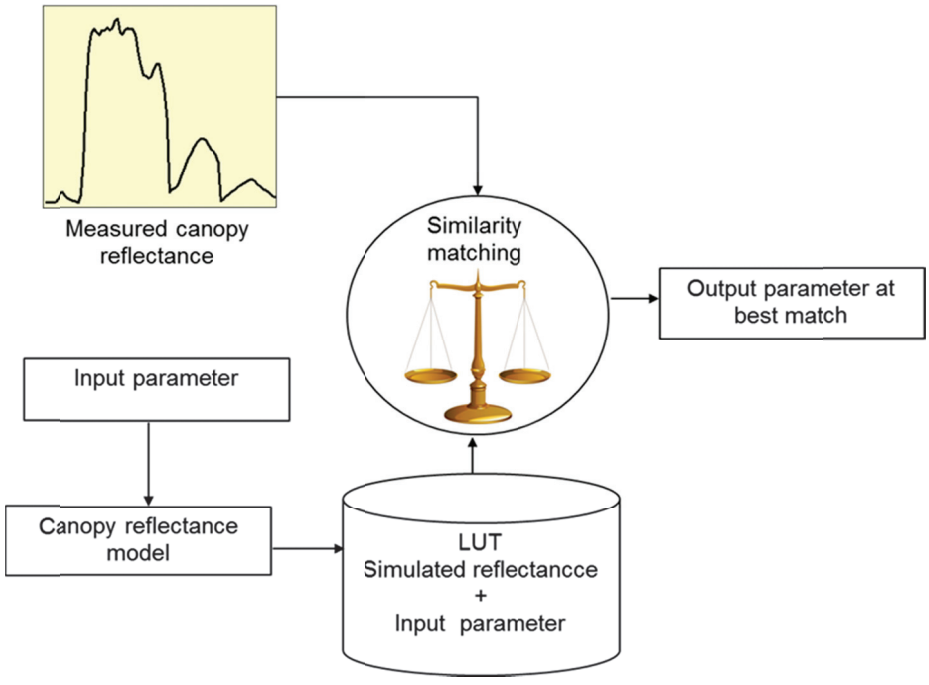

Figure 2: Conceptual diagram for general look-up-table inversion

The major advantage of the LUT-based approach is that the forward modeling is divorced from the inversion procedure, and hence can be used for any complex model like DART (Barnsley et al., 2000; Gastellu-Etchegorry et al., 2003). A key problem is to identify a suitable range and appropriate number of cases (incremental steps) for parameters to keep the size of the LUT as small as possible. The incremental steps should be fine enough to achieve a high degree of accuracy for the estimated parameters (Combal et al., 2003; Weiss et al., 2000). This requires a large number of model simulations for a single reflectance band, which is significantly cost prohibitive in the case of a 3-D model like DART and practically infeasible for hyperspectral bands.

Peddle et al. (2003) developed the multiple forward mode (MFM) technique, which is a suite of tools for LUT inversion that automates both the parameterization and search functions. Based upon the user provided or automatically-generated input parameters and ranges, sequential runs of the canopy reflectance model are stored in a database. Similarly, Hedley et al. (2009) used a method for building an adaptive look up tree (ALUT) that evenly distributes the discretization errors of tabulated reflectances in spectral space. Such tools potentially require a large number of simulations, which is not computationally feasible for DART. Peddle et al. (2007) suggested a two-stage process to initially produce a table based on wide ranges and coarse increment values to first identify the general range of model inputs that produce reflectance values similar to those of the remote sensing image. The output can then be used to identify narrower input ranges where matches occurred, for which a final table can be produced with fine increments. Such a two-stage process might help find the optimal input range for DART parameters, but does not inform users about the appropriate incremental steps.

Furthermore, not all parameter combinations within the identified range produce output similar to the image reflectance. Identification of such unrealistic combinations might significantly reduce computation time for DART.

The main objective of this study was to develop a method to build a LUT with appropriate incremental steps and realistic combinations of DART parameters to estimate LAI by model inversion using AVIRIS hyperspectral data. We initially compared DART simulations for four narrow bands to the reflectance measured by AVIRIS data to identify parameter combinations close to those occurring in nature at the time and location of image acquisition. The importance of parameters and their optimal increments were determined from a sensitivity study. Based upon the knowledge derived from the preliminary LUT, a final LUT was built for a full spectrum of AVRIS bands. We also compared two techniques, one based upon median value and the other based upon the smallest spectral angle, for their ability to accurately select a final solution from multiple closest spectra in LUT.

2. Methods

2.1 Description of the study region

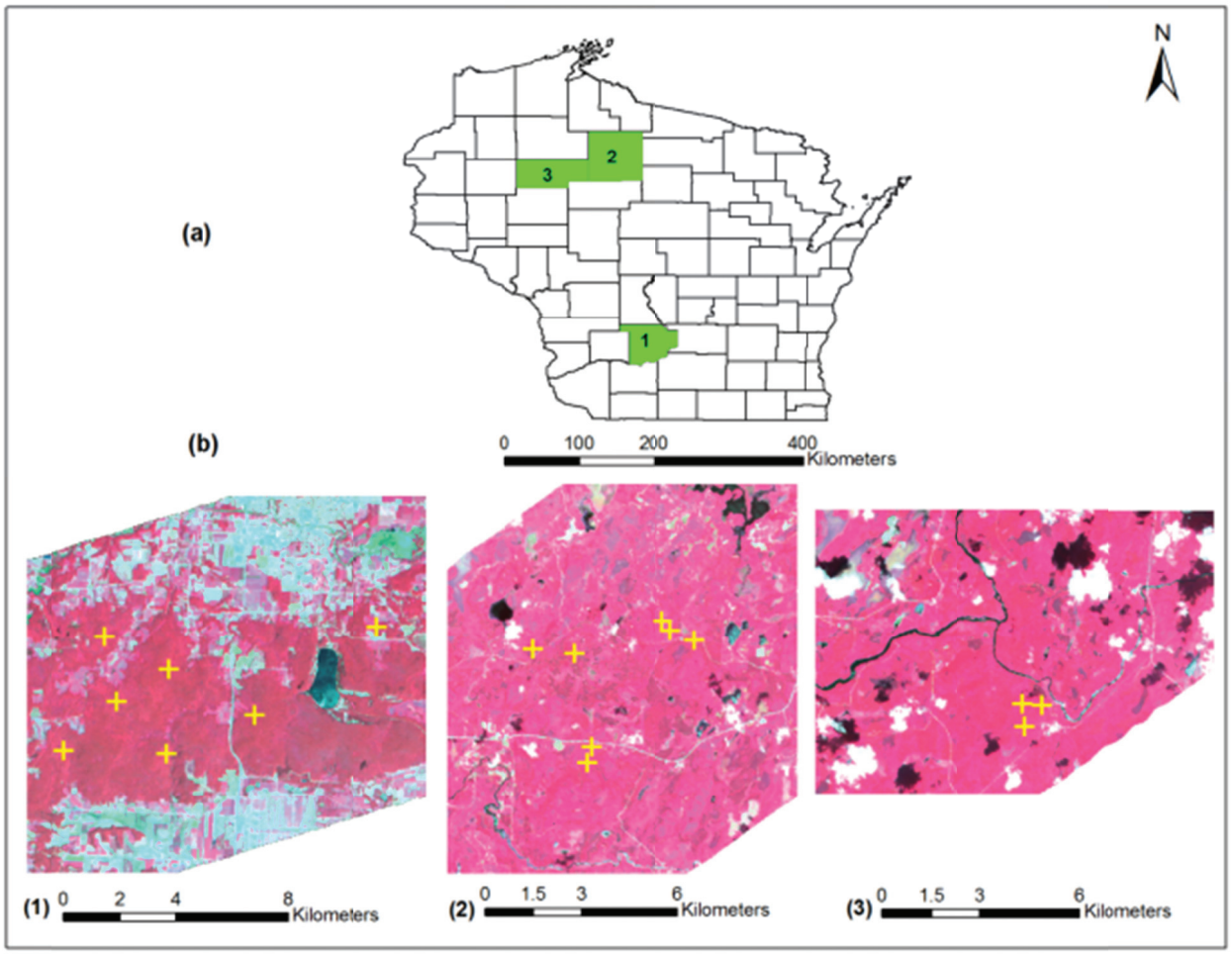

Figure 3: Study area: (a) State of Wisconsin, USA, showing general location of the three field sites and (b) plot location overlaid on a false color composite from the AVIRIS image (red: 860 nm; green: 667 nm and blue: 540 nm).

The study area comprises a range of deciduous forest types across several ecoregions within the state of Wisconsin, USA (Figure 3). The northern-most forest sites were located within the Northern Lakes and Forest ecoregion (90.120 °W, 45.936 °N to 90.318 °W, 45.966 °N), which is dominated by a mixed-hardwood forest originating from the large-scale clear-cut practices of the early 20th century (Curtis, 1959). The climate of this region is defined as northern humid continental, which is characterized by warm,

wet summers and long, cold winters (Moran and Hopkins, 2002). The overstory vegetation comprises mostly of northern hardwoods species dominated by sugar maple (*Acer saccharum*), basswood (*Tilia americana*), and white ash (*Faxinus americana*) trees.

The remaining northern forest sites were located in the highly fragmented landscape found within the Chequamegon-Nicolet National Forest, near Park Falls, Wisconsin (90.775 °W, 45.624 °N to 90.794 °W, 45.628 °N). Much of this study area is actively managed for multiple-use goals, including wood and fiber, fish and wildlife habitat, and recreation. The dominant vegetation in the upland areas comprises of northern hardwood species including sugar maple (*Acer saccharum*), basswood (*Tilia americana*), green ash (*Fraxinus pennsylvanica*), trembling aspen (*Populus tremuloides*), and paper birch (*Betula papyrifera*). The more poorly-drained, lowland deciduous sites are characterized by wetland species such as speckled alder (*Alnus incanta*) as well as hardwood species including black ash (*Fraxinus nigra*) and red maple (*Acer rubra*). The plots were positioned within the footprint of the 450 m tall WLEF broadcasting tower (45.947 °N, 90.273 °W) which contains micrometeorological instrumentation that measures the landscape- to regional-scale fluxes of CO_2, H_2O, CH_4, and energy using the eddy covariance technique (Baldocchi 2008; Cook et al. 2009; Desai et al. 2008). This tower site is a component of the Chequamegon Ecosystem-Atmosphere Study (ChEAS, http://cheas.psu.edu), which is a part of the AmeriFlux, and FLUXNET networks (Baldocchi, 2008), and serves as a NASA Earth Observing System (EOS) land validation core site (Morisette et al., 2006).

The southern sites were located in the Baraboo Hills (89.698 °W, 43.415 °N to 89.845 °W, 43.392°N) which are in the Driftless ecoregions of Wisconsin. Most of the forests in the Baraboo Hills were cleared by the 1870s and have since recovered to forests dominated by red oak (*Quercus rubra*), white oak (*Quercus alba),* green ash (*Fraxinus pennsylvanica*), white ash (*Fraxinus americana*), bitternut hickory (*Carya cordiformis*), shagbark hickory (*Carya ovata*), with sugar maple (*Acer saccharum*), red maple *(Acer rubrum*), and basswood (*Tilia americana)* in the uplands.

2.2 LAI measurement protocol

Optical measurements of effective LAI (L_e) were made at a height of one meter above the forest floor using digital hemispherical photography (DHP) (Chen et al., 2006; Jonckheere et al., 2004; Leblanc et al., 2005; Zhang et al., 2005). L_e represents the equivalent leaf area of a canopy with a random foliage distribution to produce the same light interception as the true LAI (Fernandes et al., 2004). The DHP method measures the canopy gap fraction across selected zenith angles from beneath the canopy through a hemispherical (fisheye) lens, oriented towards the zenith. For this study we utilized a Nikon CoolPix 5000 digital camera, leveled on a tripod with an attached Nikon FC-E8 fisheye lens, which has a field of view of 183°. The overstory LAI of the forest stands can be calculated using Eq.1 (Leblanc and Chen, 2001):

$$\mathrm{LAI} = \frac{Le}{\Omega} - \alpha \qquad (1)$$

where L_e is the effective LAI calculated from DHP, Ω is the clumping index and α is the woody-to-total leaf area ratio ($\alpha = W/Le\,(1/\Omega)$), where *W* represents the woody-surface-area-index (half the woody area m^{-2} ground area). In this study, we calculated

LAI from L_e by only correcting the effect of clumping and neglecting the effect of woods and branches in Eq.1 (i.e., LAI = L_e/ Ω).

DHP images were collected at 18 plots (60 m X 60 m) characterized by deciduous vegetation types spanning across the study area. All DHP measurements used in this study occurred during the peak of the summer growing seasons in 2008 and 2010 during uniformly overcast skies or during dusk or dawn when the sun was hidden by the horizon. All DHP images were collected in the JPEG format at the highest resolution (2560 x 1920 pixels) to maximize the detection of small canopy gaps (Leblanc et al., 2005). In addition, we collected images using a technique that maximizes the differentiation between leaf and sky and minimize blooming around leaves (Zhang et al., 2005). In each plot, we measured L_e at nine subplot locations: the plot center, 30 meters from the plot center in each of the four the cardinal directions, and the mid-point of each 30m transect. All DHP images were processed using the DHP software (Leblanc et al., 2005), using a nine ring configuration but selecting only the first six rings for analysis to minimize the impacts of large zenith angles on the L_e retrievals and the calculation of LAI (Chen et al., 2006; Leblanc et al., 2005). We calculated an average Ω value for each plot using the combined functionality in the DHP-TRAC software (Leblanc et al., 2005).

2.3 AVIRIS image processing

Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data used in this study (Flight ID: f080713t01 and f080714t01) were acquired in 2008 on NASA's ER-2 aircraft during the month of July (i.e., peak growing season) at an altitude of 20 km, yielding a pixel resolution of approximately 17 m (16.8 - 17.0 m). The AVIRIS instrument has 224

spectral bands, with an approximate full-width half-maximum of 10 nm for each wavelength, over the spectral range of 370 to 2500 nm (Green et al., 1998).

All AVIRIS image preprocessing involved four distinct steps: the development of an integrated cloud and cloud-shadow mask, cross-track illumination correction, removal of overlapping bands (between detectors 1 and 2 as well as between 2 and 3) and correction of atmospheric effects and conversion to top-of-canopy (TOC) reflectance. AVIRIS images often have systematic cross-track illumination effects due to a combination of flight path orientation and relative solar azimuth. We removed this effect by developing band-wise bilinear trend surfaces, ignoring all cloud/shadow-masked pixels, and trend-normalizing the images by subtracting the illumination trend surface and adding the image mean to arrive at cross-track illumination corrected bands. Atmospheric correction and conversion of the cross-track illumination corrected images to TOC reflectance was done using the ACORN5b^{TM} software (Atmospheric CORrection Now; Imspec LLC, USA). Due to the low ratio of signal to noise at both spectral ends (366 nm – 395 nm and 2467 nm – 2497 nm), and in bands around the major water absorption regions (1363 nm – 1403 nm and 1811 nm – 1968 nm), those wavelength regions were dropped resulting in a final total of 184 bands. The point map of field sampling plots was overlaid on the AVIRIS image data. The pixel reflectances corresponding to the center of the plot locations were extracted for the final 184 AVRIS bands.

Out of a total 18 deciduous plots, one with the highest LAI (6.67) had unusually low reflectance throughout the near-infrared (NIR) region (maximum reflectance of 42% at NIR plateau). We could not put our finger on what caused such low NIR reflectance in

the plot. It might be due to possible change in vegetation conditions between the time of field measurements and image acquisition or due to error in GPS coordinates. We ran a test based upon Cook's Distance which identifies influential observation on the basis of how the entire linear regression function changes when a certain observation is deleted (Cook, 1979). The test identified the plot in question as suspicious (partial F-statistic = 0.83, Cook's distance = 0.969) by its very high leverage value on regression parameter estimates. The coefficient of determination (R^2) improved to 0.72 from 0.32 after removing that plot. This outlier was thus removed from further analysis.

2.4 DART scene formulation

We used DART 4.3.3 version to simulate canopy reflectance for different parameter combinations. The DART model was run on a UNIX environment using a computer with 32 central processing units (eight Quad-Core AMD Opteron (tm) Processor 8356) and 64 GB of physical random access memory (RAM).

DART requires users to build a computer representation of a relevant earth landscape scene such as a forest, an urban area, and water body. Ideally, the scene area should be large and should include important details about the landscape. For instance, a forest scene should be built with a large number of trees, and the resolution of the scene (cell) should be high enough to represent canopy elements such as leaves, twigs, and branches. This level of details leads to unacceptable computational time requirements. Thus, a scene of reasonable size and detail needs to be determined that allows one to operate DART with an acceptable accuracy level. In this study, simulations were conducted using a simple forest representation. A repetitive forest landscape pattern made up of four trees with ellipsoid shaped crowns was chosen;

previous studies have found this representation to be optimal (Gascon et al., 2004; Gastellu-Etchegorry et al., 2003). The trees were described by slightly different heights and crown parameters. Both the scene size and the dimension of the tree crowns (major and minor axis) were varied to obtain scenes of different ground cover (see Figure 4). The canopy was represented by foliage only, without woody branches and twigs.

2.5 LUT generation

A LUT database was created using the following three steps in this study. The conceptual diagram for the LUT building process is shown in Figure 5 and discussed in the following sections.

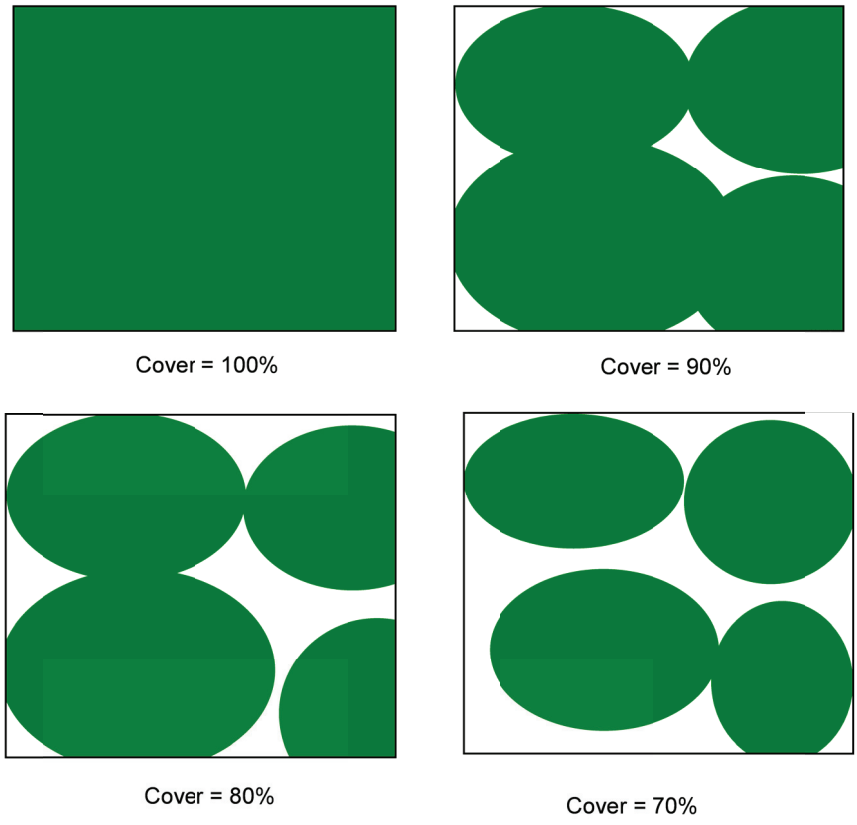

Figure 4: Tree scenes with variable tree ground covers from 100% down to 70%. DART works with an infinite scene comprised of a repetitive pattern of these tree scenes.

- A preliminary LUT was built with DART simulations for four narrow AVIRIS bands using a wide range and coarse increments for parameters.
- The simulations in LUT were compared to image reflectance, and those simulations within the range of image spectra were extracted. By analyzing the parameter values in the extracted simulations, a list of useful parameter combinations that produce reflectance values similar to AVIRIS spectra was identified. Based upon the results of a sensitivity study, optimal input increments were determined.
- A final LUT was built with DART simulations for 184 narrow AVIRIS bands using the identified list of useful parameter combinations and incremental levels.

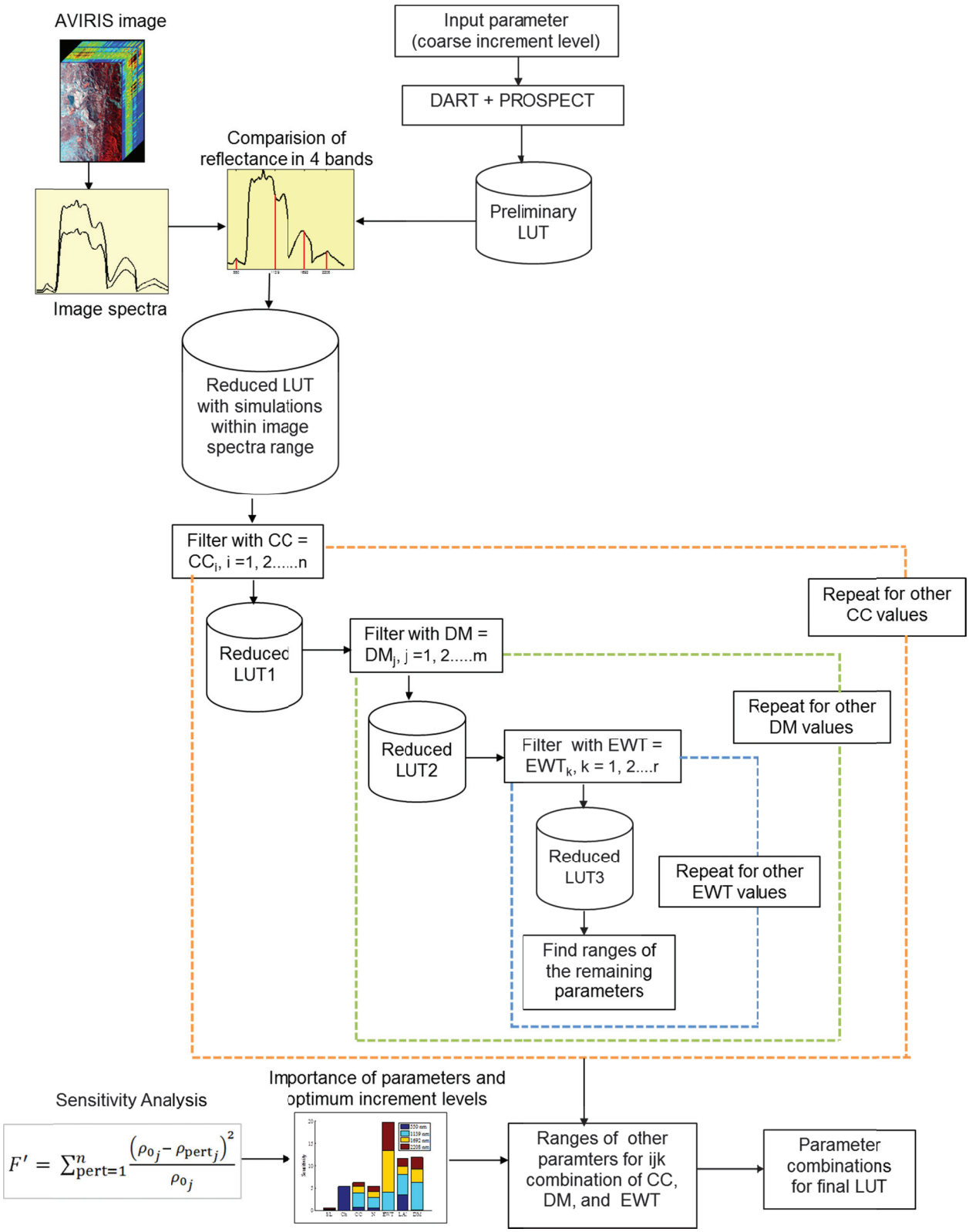

Figure 5: Conceptual framework for the LUT building technique used in this study. In the figure, CC refers to canopy cover (%), DM refers to leaf dry matter content (g/cm^{-2}), EWT refers to leaf equivalent water thickness (cm).

2.5.1 Computation of preliminary LUT

DART reflectance was simulated for the bandwidth defined by four narrow AVIRIS bands (550 nm, 1139 nm, 1692 nm, and 2208 nm). PROSPECT (Jacquemoud et al., 1993), which is part of the DART interface, was used to calculate leaf optical properties (reflectance and transmittance). PROSPECT calculates leaf hemispherical transmittance and reflectance as a function of four input parameters: leaf structural parameter (N), leaf chlorophyll a + b concentration (Cab), leaf dry matter content (DM), and equivalent water thickness (EWT). Together with the PROSPECT parameters, the parameters that were varied were leaf area index (LAI), leaf angle distribution (LAD), soil reflectance (SL), and crown cover (CC). The ranges for some of these parameters (Cab, LAI, DM, and EWT) were set based upon knowledge derived from field measurements. A fixed value for fraction of diffuse incoming solar radiation (0.1) was used across all wavelengths, as previously done in other studies (Schlerf and Atzberger, 2006). The simulations were conducted for two solar angles (SA) corresponding to the two images. The parameters, their ranges and the incremental steps are provided in Table 1. The four LADs used were Erectophile, Planophile, Plagiophile, and Extremophile distributions(de Wit, 1965).

Table 1: DART+PROSPECT parameters, their ranges, and increments used in preliminary LUT building. The parameters are leaf area index (LAI), leaf equivalent water thickness (EWT), leaf dry matter content (DM), leaf structural parameter (N), leaf chlorophyll a + b concentration (Cab), canopy cover (CC) and leaf angle distribution (LAD). The four LAD used were erectophile, planophile, plagiophile, and extremophile distributions.

Parameters	Minimum	Maximum	Increment
LAI (m^2/m^2)	2.75	6.75	0.5
EWT (cm)	0.003	0.018	0.003
DW (g/cm^2)	0.001	0.013	0.003
N (unit less)	1.5	2.5	0.5
Cab ($\mu g/cm^2$)	20	60	10
CC (%)	50	100	25
SL (%)	10	40	10
LAD	-	-	-

DART allows users to build a LUT on the basis of a sequence file in which all possible parameter values can be specified. For example, all PROSPECT and DART parameters can be written to a single sequence file and then run simultaneously to generate a single LUT. However, we modeled leaf optical properties using PROSPECT separately, and then used the subsequent output in DART to model canopy reflectance. Even though it required the creation of a large number of sequence files (one each for a single combination of PROSPECT parameters), running two models independent of each other was advantageous. Such specifications allowed up to 16 simulations in a single sequence to run in parallel, hence affording computational efficiency.

2.5.2 Search for realistic parameter combinations and sensitivity analysis

The corresponding maximum and minimum reflectance values in image extracted spectra were determined for each simulated band. The maximum and minimum values in the individual band were combined to form a maximum and a minimum reflectance (max/min) spectrum. To allow for potential model/measurement

uncertainties, the space was expanded by five percent in both directions. A search filter was applied to find the LUT simulations confined within the max/min space. The model inputs that led to simulations (candidate solutions) in the resultant reduced LUT were examined to infer information about parameter combinations with the following steps.

- Start searching candidate simulations for a single value of crown cover (for example, CC = 100%) and create a reduced-size table with only those simulations (LUT1).
- Take a single DM (for example, DM = 0.001 g/cm^2) present in LUT1 and create another reduced-size table (LUT2).
- From LUT2, create LUT3 for one of the values of EWT. Query a range of values for the remaining parameters (Cab, LAI, LAD, and N) in LUT3 and make a list of them.
- Repeat the process for all remaining values of EWT in LUT3, followed by the rest of the values of DM in LUT2 and finally all values of CC in LUT1.

The results of the above process were tabulated as optimal input ranges for four parameters for each combination of CC, DM and EWT present in candidate solutions. In principle, the process could be repeated for all combinations of parameters, but it would return a huge set of results that would be difficult to interpret. We considered CC, DM and EWT in the search process as their different values had different frequencies in candidate solutions (as revealed by their histograms, which are not shown here).All parameter values for Cab, N and SL, in contrast, were represented in about equal proportion in candidate solutions. The results would have been similar for all combinations of their values, and would not have revealed much information about optimal parameter combinations.

A simple sensitivity analysis (SA) was performed as in Privette et al. (1994) and Santis et al. (2009) to determine the importance of each parameter and the optimal discretization of parameters within their identified ranges. Each parameter was perturbed (varied) in turn keeping all other model parameters fixed at their reference values (base-case). Equal numbers of perturbation values with equal intervals were used for all parameters. The process was performed for the same set of bands previously used to compute the preliminary LUT. A merit function *F* was computed using Eq. 2 and called sensitivity:

$$F' = \sum_{\text{pert}=1}^{n} \frac{\left(\rho_{0_j} - \rho_{\text{pert}_j}\right)^2}{\rho_{0_j}} \qquad (2)$$

where, ρ_o is the base reflectance, ρ_{pert} is the perturbed reflectance in band *j*, and *n* is the total number of perturbations.

Based upon the SA results, the optimal number of cases for each parameter was determined such that the number of cases was proportional to the corresponding greatest magnitude of the sensitivity. The LAD and solar illumination angle had discrete values in this study and hence were not used in the SA method. Instead, the effects of the parameters on canopy reflectance were visually assessed by plotting the reflectance with four LADs and two solar angles simulated at constant values of the other parameters.

A final LUT was built on the basis of optimal parameter combinations and input increments determined in the previous stages. As done previously, piecewise LUTs (one table for each PROSPECT output) were built to minimize computation time. For the same reason, leaf optical properties were modeled separately for visible and

infrared bands, as reflectances in the two regions were sensitive to different PROSPECT parameters. The resultant LUTs were appended to build a final LUT.

2.5.3 LUT inversion

LUT inversion involved matching the similarity between plot spectra (measured) and simulated spectra (modeled). Spectrum matching was performed using a least root mean square error ($RMSE_r$) comparison of the measured and modeled spectra according to Eq. 3:

$$\mathrm{RMSE}_r = \sqrt{\frac{\sum_{i=1}^{n}(R_{measured} - R_{modeled})^2}{n}} \quad (3)$$

where, $R_{measured}$ is a measured reflectance at wavelength λ and $R_{modeled}$ is a modeled reflectance at wavelength λ in the LUT, and *n* is the number of wavelengths. The solution to the inverse problem was the set of input parameters corresponding to the reflectance in the database that provided the smallest $RMSE_r$. Because of potential insufficiency in model formulation and parameterization, and potential noise related to calibration and preprocessing errors in measured reflectance, the least $RMSE_r$ solution might not necessarily provide the best estimates of LAI. According to Weiss et al. (2000), the best parameter retrieval is achieved when the number of solutions ranges between 10 and 50. Consequently, for each observed spectrum, the 10, 20, 30, 40 and 50 closest matching spectra were selected from the LUT.

Out of the available multiple solutions, we used two methods to select a final solution. The first method chose the median LAI value among the multiple solutions as a final solution. In the second method, we calculated spectral angles between the

measured and the modeled spectra from multiple solutions as in spectral angle mapping classification (Kruse et al., 1993). The solution corresponding to the modeled spectra with the least angle was chosen as a final solution. The angle, α, between measured reflectance and modeled reflection was calculated using Eq. 4:

$$a = \cos^{-1} \frac{\sum_{i=1}^{n} y_i r_i}{\sqrt{\sum_{i=1}^{n} y_i^2} \sqrt{\sum_{i=1}^{n} r_i^2}} \quad (4)$$

where *y* is the modeled reflectance, *r* is the measured reflectance, and *n* is the total number of bands. Spectral angle provides a similarity measure in terms of the shape of the two compared spectra (Sohn & Rebello, 2002). Hence, we assumed that the solution which is both closest in spectral distance (as evaluated by least RMSEr) and closest in spectral shape (as evaluated by least spectral angle) might provide better estimates of LAI.

3. Results

3.1 Optimal ranges for parameters

A preliminary LUT was generated with coarse increment levels for DART and PROSPECT parameters. The resultant table was searched for simulations that were comparable to the image reflectance. Out of a total of 388,800 preliminary simulations, only 44,416 (candidate simulations) were found to be within the space bounded by the image spectra (max/min space). In candidate solutions, earlier ranges for all parameters were unchanged except for crown cover, as simulations with 50% crown cover were filtered out.

3.2 Search for realistic combinations

One of the results of the search for optimal parameter combinations is provided in Table 2. The table shows all parameter combinations in candidate simulations with 100% CC and 0.01 g/cm^2 DM that produced reflectance within the max/min space. Only three out of six EWT were present with anything equal to or less than 0.012 cm filtered out. This implied that any combinations of EWT less than or equal to 0.012 cm with 100% CC and 0.01 g/cm^2 DM were less likely to result in reflectance similar to image reflectance and could be left out at the final LUT building process. Similarly, only LAI between 4.25 and 6.75 combined with 0.015 and 0.018 EWT, 0.01 DM g/cm^2, and 100% CC resulted in candidate solutions. This suggested that we could leave out LAI values less that 4.25 and greater than 6.75 for the particular combinations of other parameters.

Table 2: One of the results from search for realistic combinations of parameters. The parameters are canopy cover (CC), leaf dry matter content (DM), leaf equivalent water thickness (EWT), leaf area index (LAI), leaf angle distribution (LAD), leaf structural parameter (N), and leaf chlorophyll a + b concentration (Cab).

CC (%)	DM (g/cm^{-2})	EWT (cm)	LAI(m^2/m^2)	LAD	N (unit less)	Cab (µg/cm^{-2})
		0.012	4.25 - 6.75	Planophile Plagiophile Extremophile	1.5, 2, 2.5	20 – 60
100	0.01	0.015	4.25 - 6.75	Planophile Plagiophile Extremophile	1.5, 2, 2.5	20 – 60
		0.018	4.25 - 6.75	Planophile	1.5, 2, 2.5	20 – 60

3.3 Sensitivity Analysis

The results of the sensitivity analysis are summarized in Figure 6. The vertical height of each colored bar represents a sensitivity value, which is a measure of the relative importance of each parameter for causing variations in reflectance over four bands. It can be seen in the figure that soil reflectance (SL) was the least important parameter across all bands. The critical parameters were:

1. Cab and LAI in the visible band (550 nm),
2. DM and LAI in the near infrared (1139 nm),
3. EWT and DM in the middle infrared (1692 nm), and
4. EWT and DM in the shortwave infrared (2208 nm)

The corresponding magnitudes of greatest sensitivities were used to decide the maximum number of parameter cases for final simulation. This resulted in four, ten, seven, five and two cases for CC, EWT, DM, Cab, and N respectively. Independent of the sensitivity study result, 16 cases were assigned for LAI, since it was the main parameter of interest. A fixed value for soil reflectance (20%) was used because of its minimal influence on reflectance across four bands. Figures 7 and 8 show the effects of four different LADs and two solar illumination angles on reflectance over four narrow bands. Plagiophile and extremophile LADs had similar effects on the reflectance in Figure 7. Similarly, Figure 8 shows little variation among the canopy reflectance at two solar angles associated with two AVIRIS images in this study. Hence, three LADs (planophile, plagiophile, and erectophile) and a single illumination angle (associated with one of the images) were used in the final simulations.

For the new cases, the realistic parameter combinations previously identified for their nearest case were considered for final simulation. For example, for four final cases for CC (70%, 80%, 90%, and 100%), the parameter combinations for CC = 75% and CC = 100% were assigned to the first and last two cases respectively. Table 3 shows the free parameters, their ranges and the number of cases used for final simulations. If we had considered all the parameter combinations, 134,400 (16 * 10 * 7 * 2 * 5 * 4 * 3) simulations for each AVIRIS band would have been required. After searching for realistic parameter combinations, the total number of final simulations was reduced to 29,800. Simulations were done separately for visible (1,920 simulations) and infrared bands (5,960 simulations) in order to reduce computation time, as reflectances in the two regions were sensitive to different PROSPECT parameters. Common parameter values in both parts of the spectrum were later used to append the two tables and make a final database with a total of 29,800 rows.

Table 3: Input ranges and number of cases for final simulations. The parameters are leaf area index (LAI), leaf equivalent water thickness (EWT), leaf dry matter content (DM), leaf structural parameter (N), leaf chlorophyll a + b concentration (Cab), canopy cover (CC) and leaf angle distribution (LAD). The three LADs used were erectophile, planophile, and plagiophile distributions.

Parameters	Minimum	Maximum	No. of cases
LAI (m^2/m^2)	2.75	6.75	16
EWT (cm)	0.003	0.018	10
DW (g/cm^{-2})	0.001	0.013	7
N (unit less)	1.75	2.25	2
Cab ($\mu g/cm^{-2}$)	20	60	5
CC (%)	70%	10%	4
LAD	-	-	3

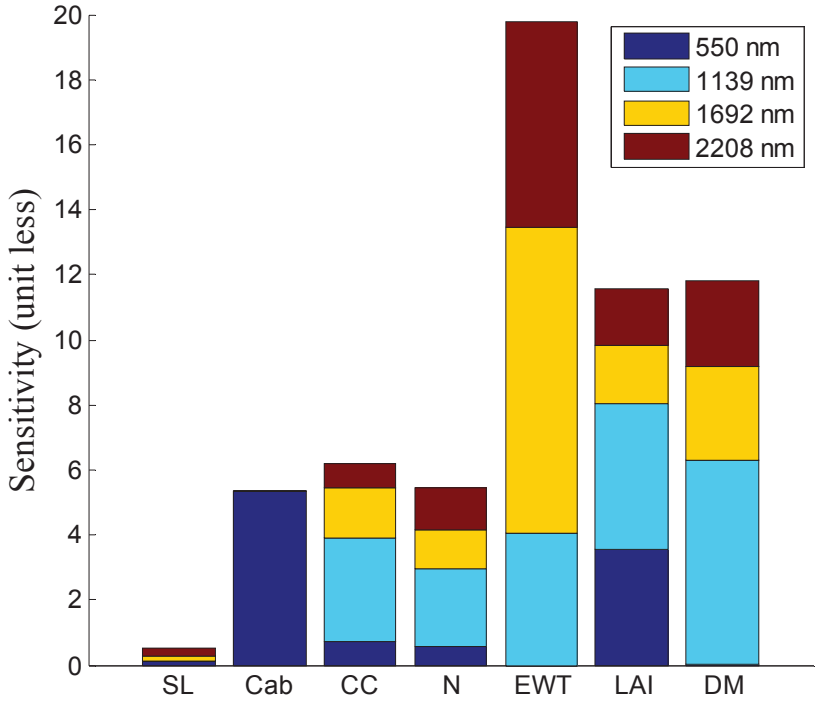

Figure 6: Results of the sensitivity analysis in four bands: 550 nm, 1139nm 1692 nm, and 2208 nm. The parameters are soil reflectance (SL), chlorophyll a and b (Cab), canopy cover (CC), leaf structure parameter (N), leaf equivalent water thickness (EWT), leaf area index (LAI), and leaf dry matter content (DM). Sensitivity refers to the relative importance of the parameters on the canopy reflectance. Each parameter was perturbed (varied) in turn keeping all other model parameters fixed at their reference values (base-case). Sensitivity was calculated as a merit function using perturbed and base-reflectance at four bands for each parameter.

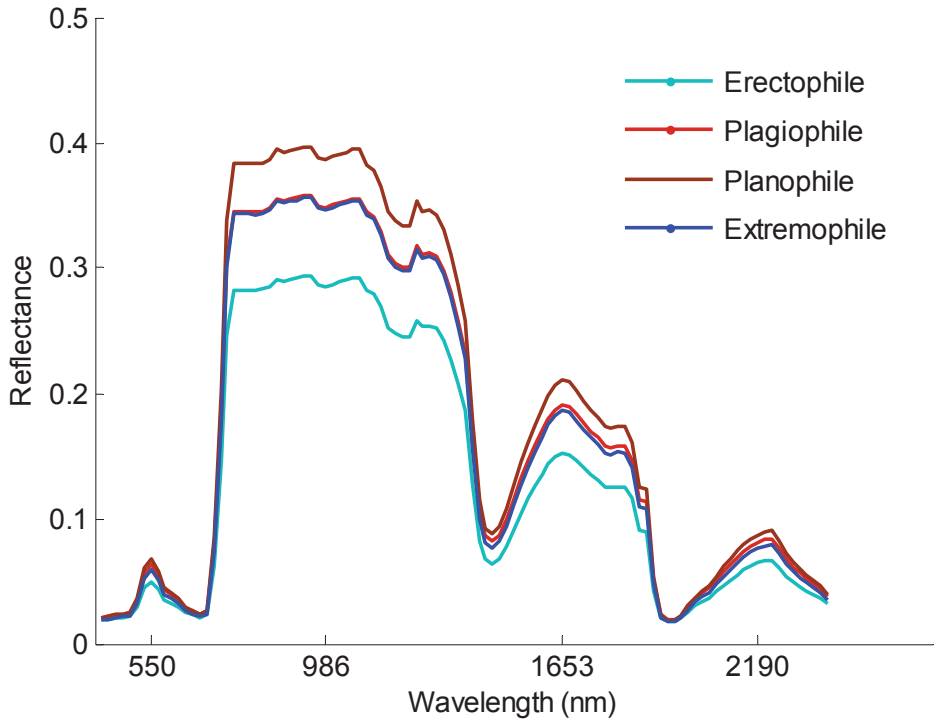

Figure 7: Effect of four different leaf angle distributions (LAD) on reflectance over AVIRIS bands at LAI = 4, CC = 85%

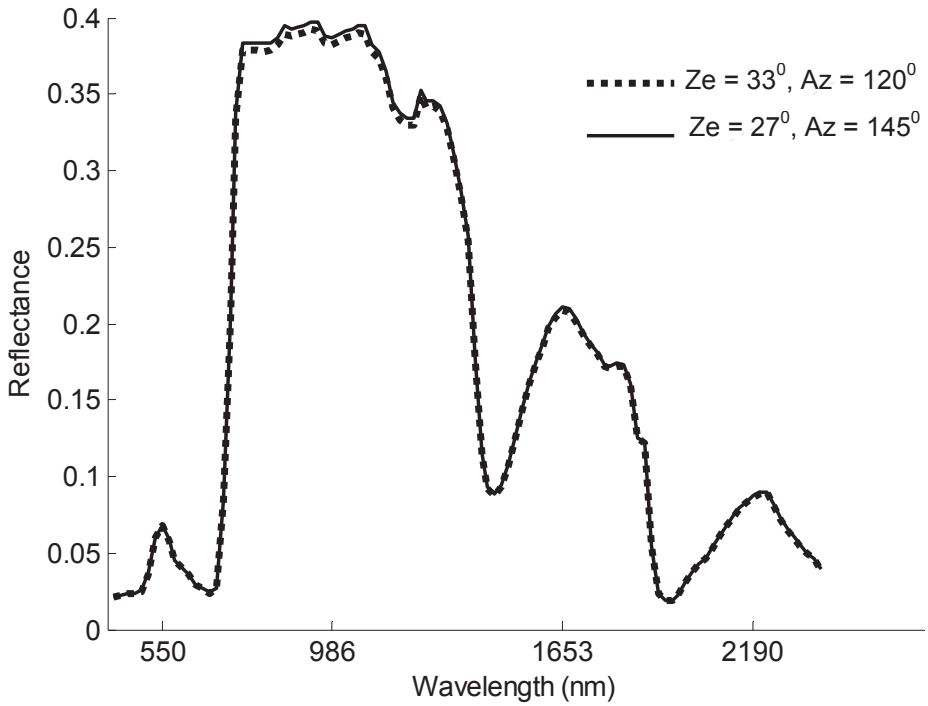

Figure 8: Effect of two different solar angles on reflectance over AVIRIS bands at LAI = 4 and CC = 85%. Ze in the figure refers to solar zenith angle and Az refers to solar azimuth angle.

3.4 Inversion

The final results from the LUT inversion are summarized in Table 4. The first column of the table shows the number of solutions considered. The second and third columns show the RMSE and R^2 between measured LAI and estimated LAI based upon median criterion. The fourth and fifth columns show the RMSE and R^2 between measured LAI and estimated LAI based upon the least spectral angle.

The results show that the LAI estimated from the median solutions had a better RMSE than the LAI estimated from the least spectral angle solutions. However, the R^2 between estimated LAI and field-observed LAI was consistently better for the least spectral angle solutions. With respect to the number of solutions, the least spectral angle calculated from 20 solutions had the lowest RMSE (Figure 9); while the median calculated from 30 solutions had the lowest RMSE for that technique (Figure 10). It can be seen that most observations were close to the 1:1 line in Figure 9, whereas all observations systematically fell above the 1:1 line in Figure 10.

Table 4: Results from the LUT inversion. The first column shows the number of solutions selected for least $RMSE_r$. The second and fourth show the RMSE and third and fifth columns show the R^2 for estimated LAI using the median and least spectral angle criteria respectively.

Number of solutions	Median solution		Least spectral angle solution	
	RMSE	R^2	RMSE	R^2
10	0.79	0.32	0.87	0.60
20	0.65	0.46	0.73	0.65
30	0.60	0.47	0.76	0.65
40	0.62	0.41	0.76	0.56
50	0.62	0.40	0.77	0.50

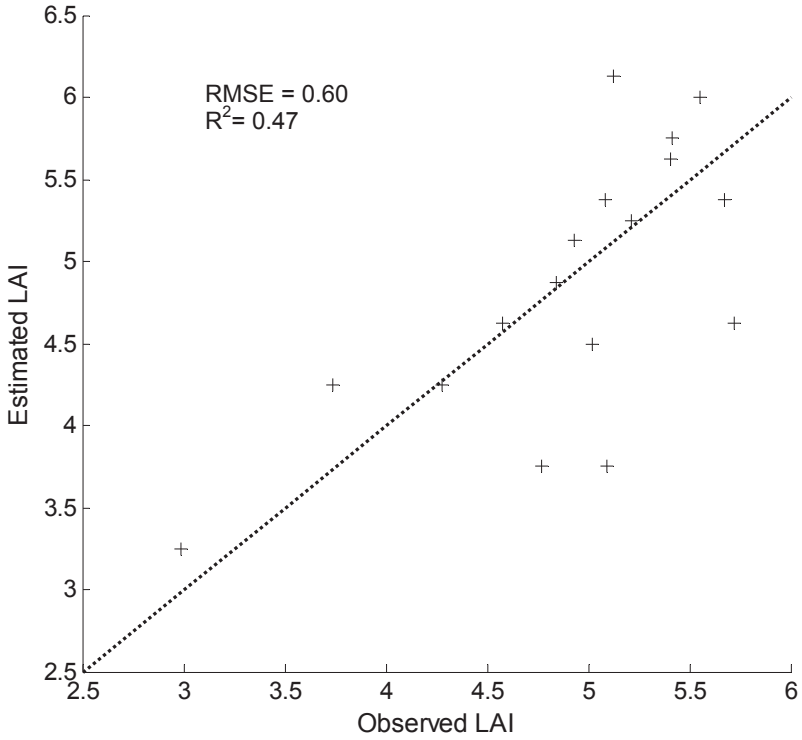

Figure 9: Observed versus estimated LAI using median value from 30 solutions

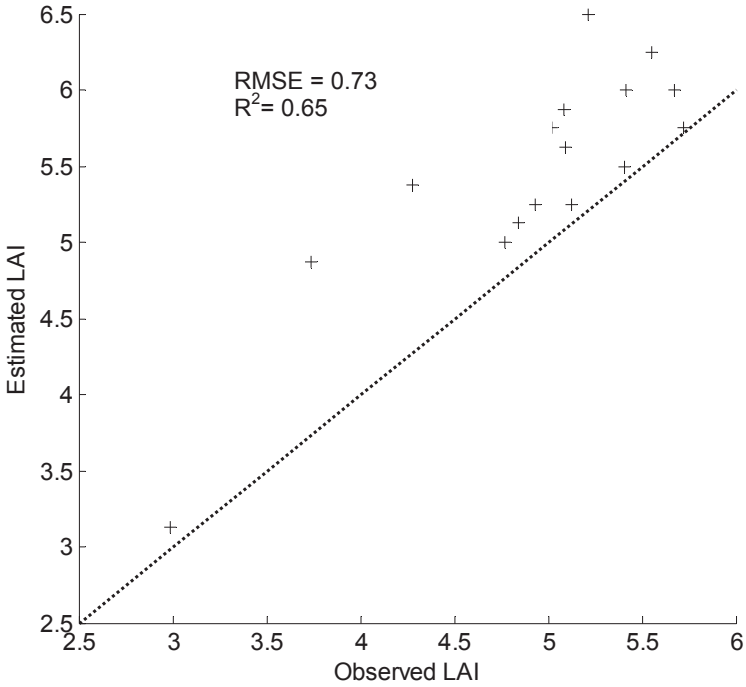

Figure 10: Observed versus estimated LAI using least spectral angle from 20 solutions

4. Discussion

In this study, we determined realistic parameter combinations and optimal input increments, enabling DART to simulate reflectance spectra commensurate to those obtained from AVIRIS. To our knowledge, no prior study has estimated canopy parameters using a complex-3D model like DART with as much spectral range and resolution. A LUT-based inversion technique using the simulated reflectance from such realistic parameter combinations resulted in reasonably accurate LAI estimates (RMSE = 0.60) in this study.

Typical DART inversions require significant computational resources. Simulating reflectance for each (AVIRIS) band increases the computation time all the more. Demarez and Gastellu-Etchegorry (2000) tested an interpolation technique for pre-computed canopy reflectance values simulated through the coupling of DART and PROSPECT. Kimes et al. (2002) also employed a similar interpolation procedure to produce directional reflectance values to estimate forest characteristics. Such interpolation procedures significantly reduce the number of simulations necessary, but still require a large number of simulations to accurately interpolate reflectance when many model inputs have unknown values. Some studies have devised automated technique of parameterization in forward modeling to build a LUT for improved inversion (Hedley et al., 2009; Peddle et al., 2003; Peddle et al., 2007). But such techniques require a model to run in multiple forward modes, which is not computationally realistic for DART. The results of this study show that the outlined three-step approach can be helpful for building a LUT of optimal size for accurate LAI estimation.

The study show that the approach can be used to identify the realistic range for an individual parameter as suggested by Peddle et al. (2007). Moreover, the results show that not all possible combinations within the identified range provide output similar to the image reflectance, and filtering out such unrealistic combinations with a simple search process results in significantly less computation time. In this study, it required about two weeks to compute reflectance for the final 28,000 simulations for AVIRIS bands using a computer with 32 processors and 64 GB of physical RAM. If we had considered all possible combinations (134,400 simulations), it would have taken nearly five times longer. The increase in computation time might not have even provided better estimates of LAI, as the additional simulations would fall outside the range of image spectra.

Sensitivity analysis helped identify both the inconsequential parameters and the number of cases for each input parameter in the final simulation. The type of sensitivity analysis used here was a simple one in the sense that it only provided information on how variation in each parameter individually produced variation in the model output. There are other methods, such as the Extended Fourier Amplitude Sensitivity Test (EFAST), which also provide information on the effect of interactions among variables in model output (Saltelli et al., 1999). Such methods, however, require a significantly greater number of simulations than required by the simple method used in this study.

The result of the inversion showed that the median based solutions had a slightly better RMSE than the least spectral angle-based solutions. The reason for including least spectral angle as an additional criterion was its perceived ability to provide a measure of similarity in shape between two spectra. Hence, it was assumed that the solution which is both closest in spectral distance (as evaluated by least RMSEr) and

closest in spectral shape (as evaluated by least angle) might provide better estimates of LAI. Figure 10 shows that the model retrieved LAI values from the least angle based solutions were in general higher than the ground-observed LAI values. This may have been due to an incorrect parameterization of understory vegetation or partial retrievals of understory LAI (e.g., Rautiainen, 2005), providing clear avenues for further study.

5. Conclusion

The main objective of this study was to determine an efficient way of building LUT for the inversion of the DART model using hyperspectral remote sensing data. We used an approach for selecting realistic parameter combinations based upon the information derived from the AVIRIS image. The approach used a coarse-resolution LUT with a limited number of bands to determine the optimal parameterization space that is likely to simulate canopy reflectance similar to that recorded by remote sensing measurements. A sensitivity study determined the optimal input increments. The results of the inversion suggested that the approach might be helpful for an efficient building of LUT to accurately estimate forest LAI by inversion of DART.

References

Asner, G.P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, 64, 234-253.

Baldocchi, D. (2008). Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany,* 56, 1-26.

Barnsley, M.J., Lewis, P., O'Dwyer, S., Disney, M.I., Hobson, P., Cutter, M., & Lobb, D. (2000). On the potential of CHRIS/PROBA for estimating vegetation canopy properties from space. *Remote Sensing Reviews*, 19, 171-189.

Chen, J.M., Govind, A., Sonnentag, O., Zhang, Y.Q., Barr, A., & Amiro, B. (2006). Leaf area index measurements at Fluxnet-Canada forest sites. *Agricultural and Forest Meteorology,* 140, 257-268.

Chopping, M., Moisen, G.G., Su, L., Laliberte, A., Rango, A., Martonchik, J.V., & Peters, D.P.C. (2008). Large area mapping of southwestern forest crown cover, canopy height, and biomass using the NASA Multiangle Imaging Spectro-Radiometer. *Remote Sensing of Environment*, 112, 2051-2063.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., & Wang, L. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, 84, 1-15.

Cook, B.D., Bolstad, P.V., Naesset, E., Anderson, R.S., Garrigues, S., Morisette, J.T., Nickeson, J., & Davis, K.J. (2009). Using LiDAR and quickbird data to model plant production and quantify uncertainties associated with wetland detection and land cover generalizations. *Remote Sensing of Environment, 113*, 2366-2379.

Cook, R.D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, 74, 169-174.

Curran, P.J. (1994). Imaging spectrometry. *Progress in Physical Geography*, 18, 247-266.

Curtis, J.T. (1959). *The vegetation of Wisconsin: An ordination of plant communities*. (2nd ed.). Madison: The University of Wisconsin Press.

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in heterogeneous grassland. *Remote Sensing of Environment*, 112, 2592-2604.

Desai, A.R., Noormets, A., Bolstad, P.V., Chen, J.Q., Cook, B.D., Davis, K.J., Euskirchen, E.S., Gough, C., Martin, J.G., Ricciuto, D.M., Schmid, H.P., Tang, J.W., & Wang, W.G. (2008). Influence of vegetation and seasonal forcing on carbon dioxide fluxes across the Upper Midwest, USA: Implications for regional scaling. *Agricultural and Forest Meteorology,* 148, 288-308.

Demarez, V., & Gastellu-Etchegorry, J.P. (2000). A modeling approach for studying forest chlorophyll content. *Remote Sensing of Environment,* 71, 226-238.

de Wit, C.T. (1965). *Photosynthesis of leaf canopies*. Wagenigen,The Netherlands: Pudoc Publ .

Duthoit, S., Demarez, V., Gastellu-Etchegorry, J.P., Martin, E., & Roujean, J.L. (2008). Assessing the effects of the clumping phenomenon on BRDF of a maize crop based on 3D numerical scenes using DART model. *Agricultural and Forest Meteorology*, 148, 1341-1352.

Fernandes, R., Miller, J.R., Chen, J.M., & Rubinstein, I.G. (2004). Evaluating image-based estimates of leaf area index in boreal conifer stands over a range of scales using high-resolution CASI imagery. *Remote Sensing of Environment*, 89, 200–216.

Gascon, F., Gastellu-Etchegorry, J.P., Lefevre-Fonollosa, M.J., & Dufrene, E. (2004). Retrieval of forest biophysical variables by inverting a 3-D radiative transfer model and using high and very high resolution imagery. *International Journal of Remote Sensing*, 25, 5601-5616.

Gastellu-Etchegorry, J.P., Demarez, V., Pinel, V., & Zagolski, F. (1996). Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote Sensing of Environment*, 58, 131-156.

Gastellu-Etchegorry, J.P., Gascon, F., & Esteve, P. (2003). An interpolation procedure for generalizing a look-up table inversion method. *Remote Sensing of Environment*, 87, 55-71.

Gastellu-Etchegorry, J.P., Martin, E., & Gascon, F. (2004). DART: a 3D model for simulating satellite images and studying surface radiation budget. *International Journal of Remote Sensing*, 25, 73-96.

Gobron, N., Pinty, B., & Verstraete, M.M. (1997). Theoretical limits to the estimation of the leaf area index on the basis of visible and near-infrared remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 1438-1445.

Gong, P., Pu, R.L., & Miller, J.R. (1995). Coniferous forest leaf-area index estimation along the Oregon transect using compact airborne spectrographic imager data. *Photogrammetric Engineering & Remote Sensing*, 61, 1107-1117.

Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M.S., Olah, M.R., & Williams, O. (1998). Imaging spectroscopy and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, 65, 227-248.

Hedley, J., Roelfsema, C., & Phinn, S.R. (2009). Efficient radiative transfer model inversion for remote sensing applications. *Remote Sensing of Environment*, 113, 2527-2532.

Huemmrich, K.F., Privette, J.L., Mukelabai, M., Myneni, R.B., & Knyazikhin, Y. (2005). Time-series validation of MODIS land biophysical products in a Kalahari woodland, Africa. *International Journal of Remote Sensing,* 26, 4381-4398.

Jacquemoud, M. (1993). Inversion of the PROSPECT + SAIL Canopy Reflectance Model from AVIRIS Equivalent Spectra: Theoretical Study. *Remote Sensing of Environment*, 44, 281-292.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., & Baret, F. (2004). Review of methods for in situ leaf area index determination - Part I. theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, 121, 19-35.

Kimes, D., Gastellu-Etchegorry, J., & Estève, P. (2002). Recovery of forest canopy characteristics through inversion of a complex 3D model. *Remote Sensing of Environment,* 79, 320-328.

Kimes, D., Knjazikhin, Y., Privette, J., Abuelgasim, A., & Gao, F. (2000). Inversion methods for physically based models. *Remote Sensing Reviews*, 18.

Kimes, D.S., Nelson, R.F., Manry, M.T. & Fung, A.K. (1998). Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *International Journal of Remote Sensing*, 19, 2639–266.

Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., & Running, S. (1998). Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research*, 103, 32257–32276.

Koetz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment*, 95,115−124.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., & Goetz, A.F.H. (1993). The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44, 145-163

Liang, S. (2007). Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography*. 31, 501-516.

Leblanc, S.G., & Chen, J.M. (2001). A practical scheme for correcting multiple scattering effects on optical LAI measurements. *Agricultural and Forest Meteorology*, 110, 125-139.

Leblanc, S.G., Chen, J.M., Fernandes, R., Deering, D.W., & Conley, A. (2005). Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agricultural and Forest Meteorology*, 129, 187-207.

Malenovski, Z., Martin, E., Homolová, L., Gastellu-Etchegorry, J.P., Zurita-Milla, R., Schaepman, M.E., Pokorný, R., Clevers, J.G.P.W., & Cudlín, P. (2008). Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution. *Remote Sensing of Environment*, 112, 1-18.

Martin, M.E., & Aber, J.D. (1997). High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological Applications*, 7, 431-443.

Mobley, C.D., Sundman, L.K., Davis, C.O., Bowles, J.H., Downes, T.V., Leathers, R.A., Montes, M.J., Bissett, W.P., Kohler, D.D.R., Reid, R.P., Louchard, E.M., & Gleason, A. (2005). Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables, *Applied Optics*, 44, 3576–3592.

Moran, J.M., & Hopkins, E.J. (2002). *Wisconsin's weather and climate*. Madison: The University of Wisconsin Press.

Morisette, J.T., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J.E., Garrigues, S., Shabanov, N.V., Weiss, M., Fernandes, R.A., Leblanc, S.G., Kalacska, M., Sanchez-Azofeifa, G.A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A.N., Lewis, T.E., Iiames, J.S., Colombo, R., Meroni, M., Busetto, L., Cohen, W.B., Turner, D.P., Warner, E.D., Petersen, G.W., Seufert, G., & Cook, R. (2006). Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1804-1817.

Myneni, R. B. (1991). Modeling radiative transfer and photosynthesis in three dimensional vegetation canopies, *Agricultural and Forest Meteorology*, 55, 323–344.

Myneni, R.B., Asrar, G., & Hall, F.G. (1992). A three-dimensional radiative transfer method for optical remote sensing of vegetated land surfaces. *Remote Sensing of Environment*, 41, 105-121, 1992.

Pastor, J., Aber, J.D., McClaugherty, C.A., & Melillo, J.M. (1984). Above-ground production and N and P cycling along a nitrogen mineralization gradient on blackhawk island, wisconsin. *Ecology*, 65, 256-268.

Peddle, D.R., Franklin, S.E., Johnson, R.L., Lavigne, M.A. & Wulder., M.A. (2003). Structural change detection in a disturbed conifer forest using a geometric optical reflectance model in multiple-forward mode. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 163-166.

Peddle, D.R., Johnson, R.L., Cihlar, J., Leblanc, S.G., Chen, J.M & Hall, F.G. (2007). Physically-based inversion modeling for unsupervised cluster labeling, independent forest classification and LAI estimation using MFM-5-scale. *Canadian Journal of Remote Sensing*, 33, 214-225.

Pinty, B., Widlowski, J.L., Taberner, M., Gobron, N., Verstraete, M.M., Disney, M., Gascon, F., Gastellu, J.P., Jiang, L., Kuusk, A., Lewis, P., Li, X., Ni-Meister, W., Nilson, T., North, P., Qin, W., Su, L., Tang, S., Thompson, R., Verhoef, W., Wang, H., Wang, J., Yan, G., & Zang, H. (2004). Radiation transfer model intercomparison (RAMI) exercise: Results from the second phase. *Journal of Geophysical Research-Atmospheres*, 109, 19.

Privette, J.L., Myneni, R.B., Tucker, C.J., & Emery, W.J. (1994). Invertibility of a 1-D discrete ordinates canopy reflectance model. *Remote Sensing of Environment*, 48, 89-105.

Rautiainen, M. (2005). Retrieval of leaf area index for a coniferous forest by inverting a forest reflectance model. *Remote Sensing of Environment*, 99, 295 -303.

Saltelli, A., Tarantola, S., & Chan, K.P.S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41, 39-56.

Santis, A.d., Chuvieco, E., & Vaughan, P.J. (2009). Short-term assessment of burn severity using the inversion of PROSPECT and GeoSail models. *Remote Sensing of Environment*, 113, 126-136.

Schlerf, M., & Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, 100, 281-294.

Schaepman, M.E., Koetz, B., Schaepman-Strub, G., & Itten, K.I. (2005). Spectrodirectional remote sensing for the improved estimation of biophysical and -chemical variables: two case studies. *International Journal of Applied Earth Observation and Geoinformation*, 6, 271-282.

Sohn, Y., & Rebello, N.S. (2002). Supervised and unsupervised spectral angle classifiers. *Photogrammetric Engineering & Remote Sensing*, 68, 1271-1280.

Verstraete, M.M., Pinty, B., & Myneni, R.B. (1996). Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing of Environment*, 58, 201-214.

Weiss, M., Baret, F., Myneni, R.B., Pragnere, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20, 3-22.

Wessman, C.A., Aber, J.D., Peterson, D.L., & Melillo, J.M. (1988). Remote-sensing of canopy chemistry and nitrogen cycling in temperate forest ecosystems. *Nature*, 335, 154-156.

Zhang, Y.Q., Chen, J.M., & Miller, J.R. (2005). Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology*, 133, 166-181.

CHAPTER 5

INVESTIGATING THE UTILITY OF WAVELET TRANSFORMS FOR INVERTING A 3-D RADIATIVE TRANSFER MODEL USING HYPERSPECTRAL DATA TO RETRIEVE FOREST LAI

Abstract

The need for a standard technique for optimal spectral sampling of hyperspectral data during the inversion of canopy reflectance models has been the subject of many studies. The objective of this study was to investigate the utility of the discrete wavelet transform (DWT) for extracting useful features from hyperspectral data with which forest LAI can be estimated through inversion of a three dimensional radiative transfer model, DART. DART, coupled with the leaf optical properties model PROSPECT, was inverted with AVIRIS data using a look-up-table (LUT)-based inversion approach. We used AVRIS data and *in situ* LAI measurements from two different hardwood forested sites in Wisconsin, USA. Prior to inversion, model-simulated and AVIRIS hyperspectral data were transformed into discrete wavelet coefficients using Haar wavelets. The LUT inversion was performed with three different datasets, the original reflectance bands, the full set of wavelet extracted features, and two wavelet subsets containing 99.99% and 99.0% of the cumulative energy of the original signal. The energy subset containing 99.99% of the cumulative signal energy provided better estimates of LAI (RMSE = 0.46, R^2 = 0.77) than the original spectral bands (RMSE = 0.60, R^2 = 0.47). The results indicate that the discrete wavelet transform can increase the accuracy of LAI estimates by improving the LUT-based inversion of DART (and, potentially, by implication, other terrestrial radiative transfer models) using hyperspectral data.

Keywords: *LAI, hyperspectral, AVIRIS, DART, LUT, inversion, discrete wavelet transform*

1. Introduction

Leaf area index (LAI) is a basic vegetation canopy property that controls and moderates different climatic and ecological functions (Huemmrich et al., 2005; Malenovski et al., 2008). Because LAI directly affects canopy spectral reflectance, remote sensing techniques offer a practical means of estimating LAI at the landscape level (Gong et al., 1995; Koetz et al., 2005; Schaepman et al., 2005). The majority of studies estimating LAI have employed empirically-based techniques, which are site-specific and may not be extendable to operational use, given different structural and climatic conditions (Curran, 1994; Gobron et al., 1997). On the other hand, the physical approach using canopy reflectance models is based upon an understanding of the physical laws governing the transfer of solar radiation in vegetative canopies; it works with variable measurement conditions and is better suited for many large scale applications (Gobron et al., 1997; Kimes et al., 2000; Verstraete et al., 1996). Physically-based methods can also make full use of the high dimensional spectral and multiangular information provided by modern sensors (Chopping et al., 2008; Darvishzadeh et al., 2008).

The majority of physically-based models are one-dimensional (1-D) in the sense that canopies vary only with the height above the ground surface. However, most plant stands contain partial cover and generally exhibit horizontal variability in their structural and optical properties, and are thus poorly represented by 1-D models. The most realistic description of reflected radiation from a forest canopy can be provided by 3-D

models (Gastellu-Etchegorry et al., 1996; Gastellu-Etchegorry et al., 2003; Kimes et al., 2000; Malenovski et al., 2008; Myneni et al., 1991; Myneni, 1992). The models are referred to as 3-D because the extinction and scattering coefficients that define photon interactions are an explicit functions of the spatial coordinates (Myneni, 1992).

DART is a widely used 3-D model in several vegetation-related scientific works (Duthoit, 2008; Gascon et al., 2004; Gastellu-Etchegorry et al., 2004; Malenovski et al., 2008). Though it has a large set of input parameters, the number of parameters can be greatly reduced through simple approximation, making it amenable to inversion to estimate model parameters (Gastellu-Etchegorry et al., 1996). Since its first release (Gastellu-Etchegorry et al., 1996), its accuracy, range of applications, and graphical user interface have been significantly improved (Gascon et al., 2001; Malenovski et al., 2008). DART uses simplifying hypotheses for simulating vegetation landscapes by dividing the scene into rectangular cells characterized by the volume and scattering properties of landscape elements. It can accurately model multiple scattering effects within canopies (Gastellu-Etchegorry et al., 2004). Radiative scattering and propagation are simulated with the exact kernel and discrete ordinate approaches. The model output predicts any specified directional sensor response. It was compared and tested within the radiation transfer model intercomparison (RAMI) experiment (Pinty et al., 2004) and had an unprecedented level of agreement with other candidate models in simulating heterogeneous canopy spectra (Duthoit et al., 2008). Figure 1 shows an example of DART computer representation of landscape elements. A detailed description about the DART model can be found in Gastellu-Etchegorry et al. (1996) and Gastellu-Etchegorry et al. (2003).

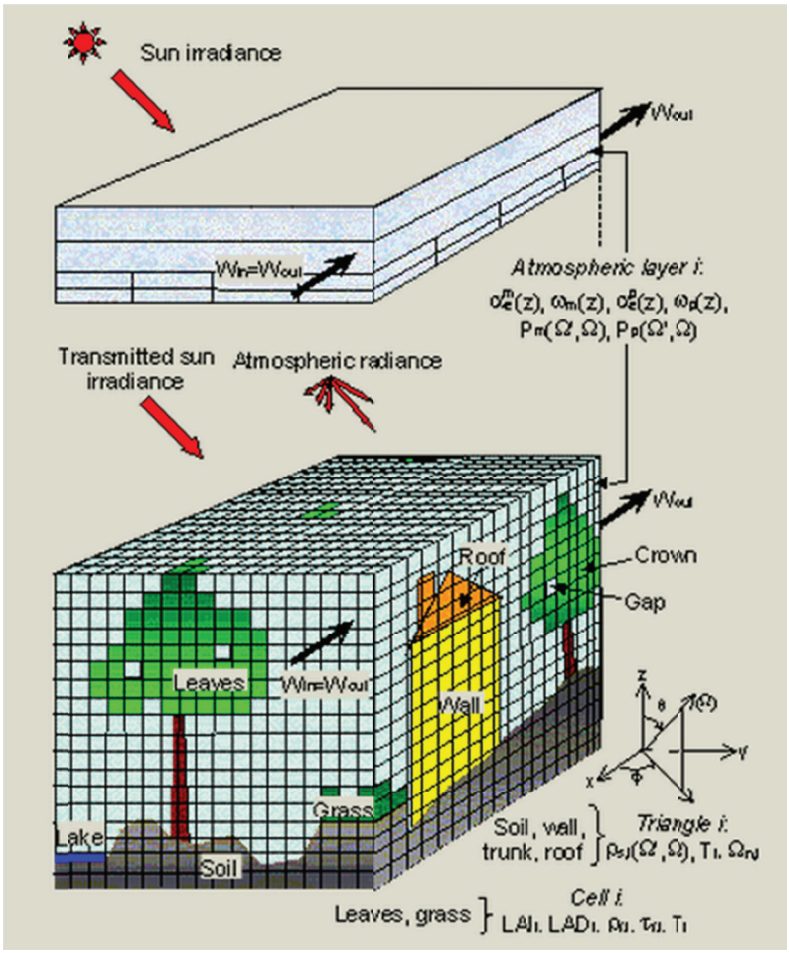

Figure 1: An example of DART computer representation of landscape elements, atmospheric layer, and illumination. The different symbols represent variables that define the optical properties of each DART cell. (Source:http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI3/MODELS/DART/DART.php)

For the estimation of vegetation properties such as LAI, a model has to be inverted against measured reflectance data. Several model inversion techniques are available (Kimes et al., 1998; Kimes et al., 2000; Schlerf and Atzberger, 2006). Traditional inversion methods like optimization techniques iteratively adjust model parameters until the modeled reflectance 'fits' or matches measured reflectance. These iterative optimization approaches are not feasible for complex models with slow computation time when simulating reflectance over a large number of spectral bands. Recently, the LUT-based inversion methods have been widely used (Darvishzadeh et al., 2008;

Hedley et al., 2009; Knyazikhin et al., 1998; Liang, 2007). They require construction of a database of pre-calculated reflectances as a function of key model parameters. Then, the spectrum for a particular image pixel is compared with each spectrum in the database, and the closest match to the image spectrum is found. The parameter combination that yields the closest spectrum in the database is considered to be the inversion solution (Mobley et al., 2005). The major advantage of the LUT-based approaches is that the forward modeling is divorced from the inversion procedure, and hence can be used for inversion of any complex model like DART (Barnsley et al., 2000; Gastellu-Etchegorry et al., 2003). The conceptual diagram for a general LUT-based inversion is shown in Figure 2.

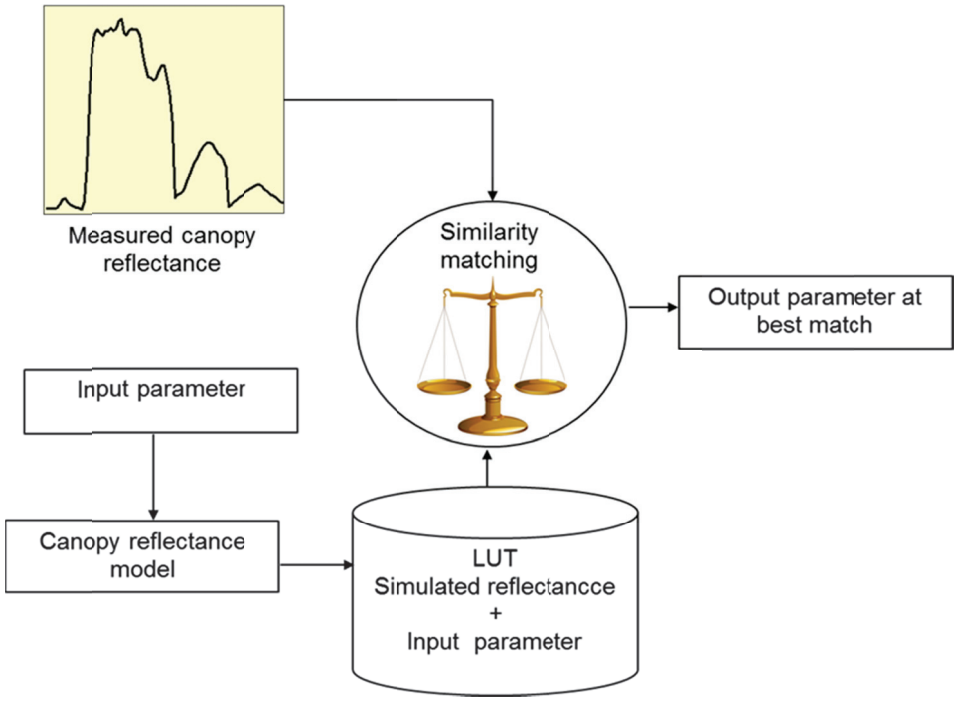

Figure 2: Conceptual diagram for general look-up-table inversion

Model inversion requires that for a unique determination of *M* parameters from *N* independent measurements (directional or spectral remote sensing measurements), *M*

must be less than or equal to *N* (Goel, 1988; Kimes et al., 2000; Verstraete et al., 1996). Hyperspectral sensors subdivide the optical portion of the electromagnetic spectrum into a large number of narrow spectral bands. However, the inherent dimensionality of hyperspectral bands could be much lower than the total number of bands because of the noisier data and greater correlation in the measurements (Schaepman et al., 2009). Such highly correlated bands artificially carry exaggerated weight and cause inversion bias (Bacour et al., 2002; Meroni et al., 2004). Weiss et al. (2000) reported that only a limited number of wavebands are required for canopy biophysical variable estimation. Other studies have stated that the selection of a subset of spectral bands can lead to a more stable and accurate inversion (Barnsley et al., 2000; Meroni et al., 2004). However, a general criterion for the selection of bands has not yet been defined. Weiss et al. (2000) and Meroni et al. (2004) utilized an empirical technique in which, starting from a combination of three bands, they selected bands in a stepwise manner. Such a stepwise procedure is a sub-optimal search as there is no guarantee of finding the best set of all possible combinations (Meroni et al., 2004; Weiss et al., 2000).

The discrete wavelet transform (DWT) has been increasingly used in recent years in the processing of hyperspectral image for a variety of purposes (Banskota et al., 2011; Blackburn, 2007; Blackburn and Ferwerda, 2008; Bradshaw and Spies, 1992; Bruce et al., 2001; Kaewpijit et al., 2003; Pu and Gong, 2004; Zhang et al,. 2006). These studies have shown its importance as an effective tool for reducing the dimensionality of hyperspectral data while maintaining information content for a variety of applications. The utility of the wavelet transform for model inversion has also been explored and confirmed in several geophysical and seismic studies, where it is integrated in the

forward modeling process and the model is inverted in wavelet domain (Ji et al., 2002; Miller and Willsky, 1996). Such direct integration of DWT in forward modeling is also a possibility for canopy reflectance models. Alternatively, the transformation and inversion can be done in a completely offline mode (independent of forward modeling) by transforming the LUT entries and remote sensing measurements into wavelet features.

The advantages of DWT in LUT inversion are threefold. First, it transforms the correlated set of spectral features into wavelet features at different scales; the information carried at different scales are less correlated because of the orthogonal nature of transformation (Zhang et al., 2006). Second, most of the wavelet coefficients contain very low values which can be discarded by applying an appropriate threshold without losing significant information, thus affording dimensionality reduction. Third and most importantly, the wavelet transform affords a multi-resolution analysis of hyperspectral data by decomposing it into components of different spectral scales (analogous to spectral bands of different width). The sensitivity of canopy reflectance is scale dependent, as some parameters affect reflectance locally in a narrow range of spectral bands (e.g., chlorophyll), whereas, parameters like LAI, canopy cover and leaf angle affect reflectance over a much wider region of the spectrum. The multi-resolution analysis by discrete wavelet transform offers the possibility of re-distributing the effects of different parameters on canopy reflectance in a single band over wavelet coefficients at multiple scales. Hence, an individual wavelet coefficient might be sensitive to a reduced set of parameters, in contrast to the corresponding spectral measurement, potentially leading to an improved inversion of parameters.

The main objective of this study was to assess the utility of the Haar discrete wavelet transform as a feature extraction technique for accurately estimating forests LAI through LUT inversion of DART using airborne hyperspectral data (AVIRIS). We posited that the transformation of hyperspectral reflectance into wavelet features and subsequent LUT inversion in the wavelet domain might provide improved estimates of LAI. We also examined three different sets of wavelet features and their effect on inversion accuracy.

2. Background on wavelet transforms

The following brief introduction about wavelet transforms is taken from Banskota et al. (2011). More detailed discussion can be found elsewhere (e.g. Graps, 1995).

An oscillating function with a limited time duration that is irregular, asymmetric, and has an average value of zero can be regarded as a wavelet function. A wavelet transform enables signal (data) analysis at different scales or resolutions by creating a series of shifted and scaled versions of the mother wavelet function. The term “mother” implies that a set of basis functions $\{\psi_{a,b}(\lambda)\}$ can be generated from one main function, or the mother wavelet $\psi(\lambda)$ by the following equation (Bruce et al., 2001),

$$\psi_{a,b}(\lambda) = \frac{1}{\sqrt{a}}\psi\left(\frac{\lambda - b}{a}\right) \quad (1)$$

where, a is the scaling factor of a particular basis function and b is the translation variable along the function’s range.

Wavelet analysis is based upon the idea of projecting a signal onto the basis functions to resolve the signal at different scales or resolutions (Hsu, 2007). Resolving data into different scales is analogous to looking at data with windows of different width. The ‘gross’ features of data might be visible at large scales and ‘fine’ features at small scales (Banskota et al., 2011; Blackburn, 2007). The results of wavelet analysis of a signal are wavelet coefficients, which are a function of the scale of the analyzing wavelet and the position of the signal (part of the signal being analyzed). Calculating wavelet coefficients at every possible scale and position is computationally intensive and generates enormous amounts of data. In the discrete wavelet transform (DWT), wavelet coefficients are usually sampled at some discrete scale and positions based on powers of two (dyadic scales and positions).

An efficient way to implement the DWT scheme using filters was developed by Mallat (1989). The basic idea behind this fast algorithm is to represent the wavelet basis as a set of high-pass and low-pass filters. The high pass and low pass filters are related (their power sum is equal to one) and called quadrature mirror filters. Thus, the first level of DWT decomposition of a signal splits it into a low pass version and a high pass version. The second level of decomposition is performed on the low pass signal obtained from the first level of decomposition. The wavelet decomposition can be iteratively performed until a maximum scale is reached. The maximum scale is dependent on the signal length and the wavelet basis length. The final results of a DWT decomposition of a spectrum are sets of wavelet coefficients, with each wavelet coefficient directly related to the amount of energy in the signal at different positions and scales.

3. Methods

3.1 Description of the study region

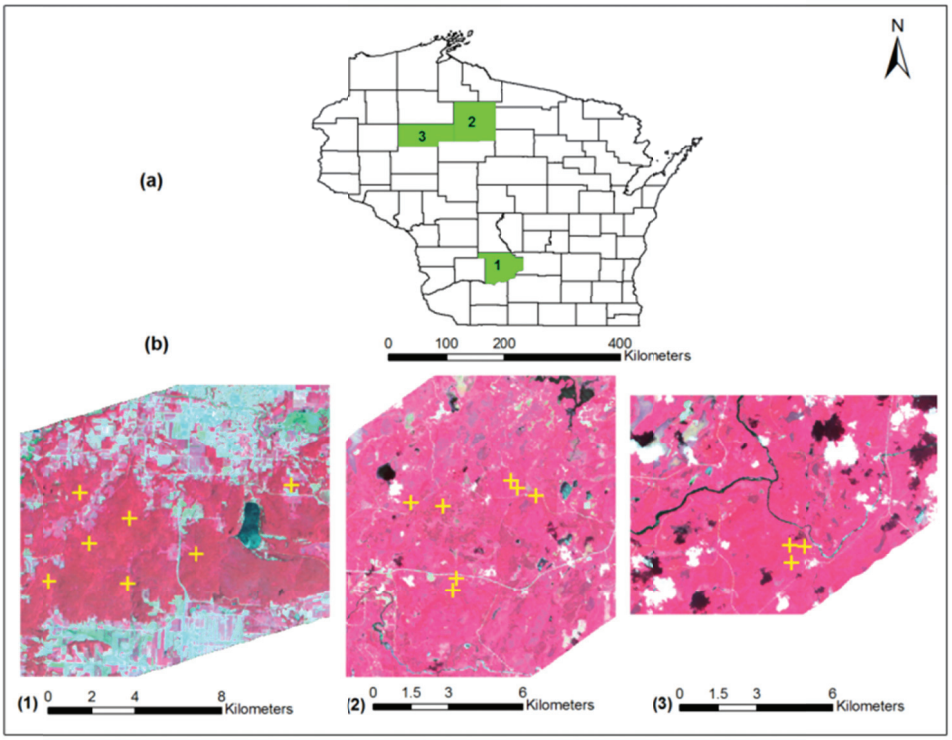

Figure 3: Study area: (a) State of Wisconsin, USA, showing general location of the three field sites and (b) plot location overlaid on a false color composite from the AVIRIS image (red: 860 nm; green: 667 nm and blue: 540 nm)

The study area comprises a range of deciduous forest types across several ecoregions within the state of Wisconsin, USA (Figure 3). The northern-most forest sites were located within the Northern Lakes and Forest ecoregion (90.120 °W, 45.936 °N to 90.318 °W, 45.966 °N), which is dominated by a mixed-hardwood forest originating from the large-scale clear-cut practices of the early 20th century (Curtis, 1959). The climate of this region is defined as northern humid continental, which is characterized by warm, wet summers and long, cold winters (Moran and Hopkins, 2002). The overstory vegetation is comprised mostly of northern hardwoods dominated by sugar maple (*Acer saccharum*), basswood (*Tilia americana*), and white ash (*Faxinus americana*) trees.

The remaining northern forest sites were located in the highly fragmented landscape found within the Chequamegon-Nicolet National Forest, near Park Falls, Wisconsin (90.775 °W, 45.624 °N to 90.794 °W, 45.628 °N). Much of this study area is actively managed for multiple-use goals, including wood and fiber, fish and wildlife habitat, and recreation. The dominant vegetation in the upland areas comprises of northern hardwood species including sugar maple (*Acer saccharum*), basswood (*Tilia americana*), green ash (*Fraxinus pennsylvanica*), trembling aspen (*Populus tremuloides*), and paper birch (*Betula papyrifera*). The more poorly-drained, lowland deciduous sites were characterized by wetland species such as speckled alder (*Alnus incanta*) as well as hardwood species including black ash (*Fraxinus nigra*) and red maple (*Acer rubra*). The plots were positioned within the footprint of the 450 m tall WLEF broadcasting tower (45.947 °N, 90.273 °W) which contains micrometeorological instrumentation that measures the landscape- to regional-scale fluxes of CO_2, H_2O, CH_4, and energy using the eddy covariance technique (Baldocchi, 2008; Cook et al., 2009; Desai et al., 2008). This tower site is a component of the Chequamegon Ecosystem-Atmosphere Study (ChEAS, http://cheas.psu.edu), which is a part of the AmeriFlux, and FLUXNET networks (Baldocchi, 2008), and serves as a NASA Earth Observing System (EOS) land validation core site (Morisette et al., 2006).

The southern sites were located in the Baraboo Hills (89.698 °W, 43.415 °N to -89.845 °W, 43.392°N) which are in the Driftless ecoregions of Wisconsin. Most of the forests in the Baraboo Hills were cleared by the 1870s and have since recovered to forests dominated by red oak (*Quercus rubra*), white oak (*Quercus alba),* green ash (*Fraxinus pennsylvanica*), white ash (*Fraxinus americana*), bitternut hickory (*Carya*

cordiformis), shagbark hickory (*Carya ovata*), with sugar maple (*Acer saccharum*), red maple *(Acer rubrum*), and basswood (*Tilia americana)* in the uplands.

3.2 LAI measurement protocol

Optical measurements of effective LAI (L_e) were made at a height of one meter above the forest floor using digital hemispherical photography (DHP) (Chen et al., 2006; Jonckheere et al., 2004; Leblanc et al., 2005; Zhang et al., 2005). L_e represents the equivalent leaf area of a canopy with a random foliage distribution to produce the same light interception as the true LAI (Fernandes et al., 2004). The DHP method measures the canopy gap fraction across selected zenith angles from beneath the canopy through a hemispherical (fisheye) lens, oriented towards the zenith. For this study we utilized a Nikon CoolPix 5000 digital camera, leveled on a tripod with an attached Nikon FC-E8 fisheye lens, which has a field of view of 183°. The overstory LAI of the forest stands can be calculated using Eq.2 (Leblanc and Chen, 2001):

$$\mathrm{LAI} = \frac{Le}{\Omega} - \alpha \qquad (2)$$

where L_e is the effective LAI calculated from DHP, Ω is the clumping index and α is the woody-to-total leaf area ratio ($\alpha = W/Le\,(1/\Omega)$) , where *W* represents the woody-surface-area-index (half the woody area m^{-2} ground area). In this study, we calculated LAI from L_e by only correcting the effect of clumping and neglecting the effect of woods and branches in Eq. 2 (i.e., LAI = L_e/ Ω).

DHP images were collected at 18 plots (60 m X 60 m) characterized by deciduous vegetation types spanning across the study area. All DHP measurements used in this study occurred during the peak of the summer growing seasons in 2008

and 2010 during uniformly overcast skies or during dusk or dawn when the sun was hidden by the horizon. All DHP images were collected in the JPEG format at the highest resolution (2560 x 1920 pixels) to maximize the detection of small canopy gaps (Leblanc et al., 2005). In addition, we collected images using a technique that maximizes the differentiation between leaf and sky and minimize blooming around leaves (Zhang et al., 2005). In each plot, we measured L_e at nine subplot locations: at the plot center, 30 meters from the plot center in each of the four cardinal directions, and at the mid-point of each 30 meter transect. All DHP images were processed using the DHP software (Leblanc et al., 2005), using a nine ring configuration but selecting only the first six rings for analysis to minimize the impacts of large zenith angles on the L_e retrievals and the calculation of LAI (Chen et al., 2006; Leblanc et al., 2005). We calculated an average Ω value for each plot using the combined functionality in the DHP-TRAC software (Leblanc et al., 2005).

3.3 AVIRIS image processing

Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data used in this study (Flight ID: f080713t01 and f080714t01) were acquired in 2008 on NASA's ER-2 aircraft during the month of July (i.e., peak growing season) at an altitude of 20 km, yielding a pixel (i.e., spatial) resolution of approximately 17 m (16.8 - 17.0 m). The AVIRIS instrument produces 224 spectral bands (or wavelengths), with an approximate full-width half-maximum of 10 nm for each wavelength, over the spectral range of 370 to 2500 nm (Green et al. 1998).

All AVIRIS image preprocessing involved four distinct steps: the development of an integrated cloud and cloud-shadow mask, cross-track illumination correction, removal of

overlapping bands (between detectors 1 and 2 as well as between 2 and 3) and correction of atmospheric effects and conversion to top-of-canopy (TOC) reflectance. AVIRIS images often have systematic cross-track illumination effects due to a combination of flight path orientation and relative solar azimuth. We removed this effect by developing band-wise bilinear trend surfaces, ignoring all cloud/shadow-masked pixels, and trend-normalizing the images by subtracting the illumination trend surface and adding the image mean to arrive at cross-track illumination corrected bands. Atmospheric correction and conversion of the cross-track illumination corrected images to TOC reflectance was done using the ACORN5b^{TM} software (Atmospheric CORrection Now; Imspec LLC, USA). Due to the low ratio of signal to noise at both spectral ends (366 nm – 395 nm and 2467 nm – 2497 nm), and in bands around the major water absorption regions(1363 nm – 1403 nm and 1811 nm – 1968 nm), those wavelength regions were ommitted resulting in a total of 184 number of bands. The point map of field sampling plots was overlaid on the AVIRIS image data. The pixel reflectances corresponding to the center of the plot locations were extracted for the final 184 AVRIS bands.

Out of a total 18 deciduous plots, one with the highest LAI (6.67) had unusually low reflectance throughout the near-infrared (NIR) region (maximum reflectance of 42% at NIR plateau). We could not put our finger on what caused such low NIR reflectance in the plot. It might be due to possible change in vegetation conditions between the time of field measurements and image acquisition or due to error in GPS coordinates. We ran a test based upon Cook's Distance which identifies influential observation on the basis of how entire linear regression function changes when a certain observation is deleted

(Cook, 1979). The test identified the plot in question as suspicious (partial F-statistic = 0.83, Cook's distance = 0.969) by its very high leverage value on regression parameter estimates. The coefficient of determination (R^2) improved to 0.72 from 0.32 after removing that plot. This outlier was thus removed from further analysis.

3.4 DART scene formulation

We used DART 4.3.3 version to simulate canopy reflectance for different parameter combinations. The DART model was run on a UNIX environment using a computer with 32 processors (eight Quad-Core AMD Opteron (tm) Processor 8356) and 64 GB of physical random access memory (RAM).

DART requires users to build a computer representation of a relevant earth landscape scene such as a forest, an urban area, and water body. Ideally, the scene area should be large and should include important details about the landscape. For instance, a forest scene should be built with a large number of trees, and the resolution of the scene (cell) should be high enough to represent canopy elements such as leaves, twigs, and branches. This level of details leads to unacceptable computational time requirements. Thus, a scene of reasonable size and detail needs to be determined that allows one to operate DART with an acceptable accuracy level. In this study, simulations were conducted using a simple forest representation. A repetitive forest landscape pattern made up of four trees with ellipsoid shaped crowns was chosen; previous studies have found this representations to be optimal (Gascon et al., 2004; Gastellu-Etchegorry et al., 2003). The trees were described by slightly different heights and crown parameters. Both the scene size and the dimension of the tree crowns (major and minor axis) were varied to obtain scenes of different ground cover (see

Figure 4). The canopy was represented by foliage only, without woody branches and twigs.

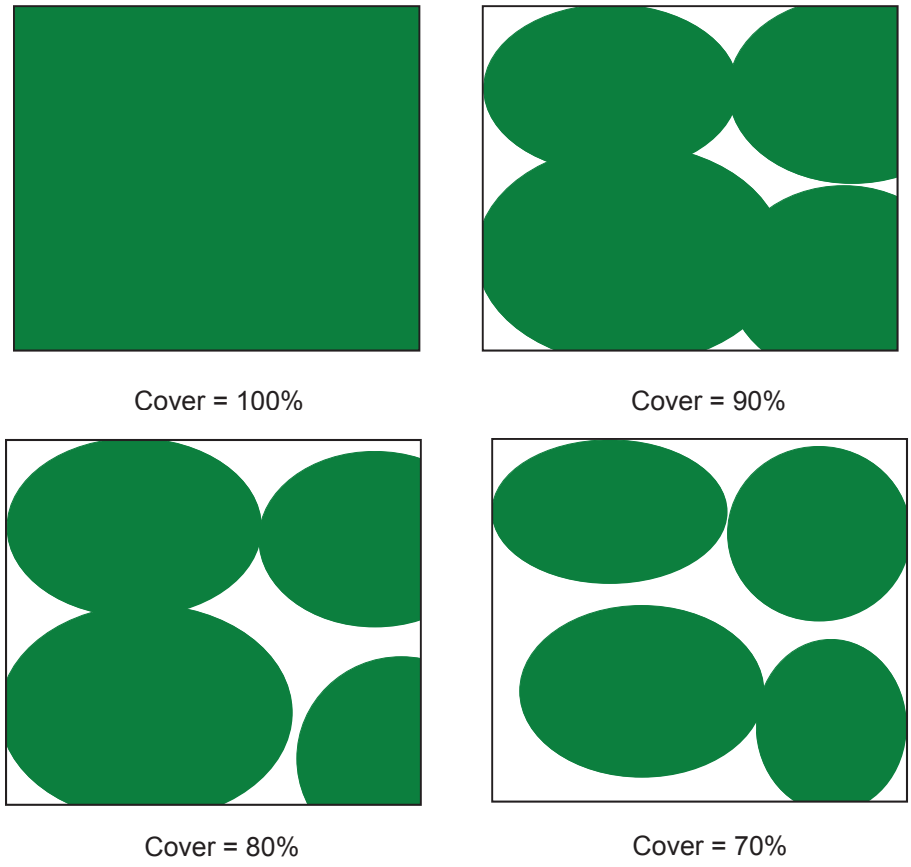

Figure 4: Tree scenes with variable tree ground covers from 100% down to 70%. DART works with an infinite scene comprised of a repetitive pattern of these tree scenes.

3.5 Creation of LUT database

Construction of a LUT of reflectance values for a large number of bands with a 3-D model such as DART requires an unacceptable computational time. In this study, we used the same set of LUT database as used by Banskota et al. (manuscript in preparation). The conceptual diagram for the LUT building process is shown in Figure 5 and is briefly discussed below.

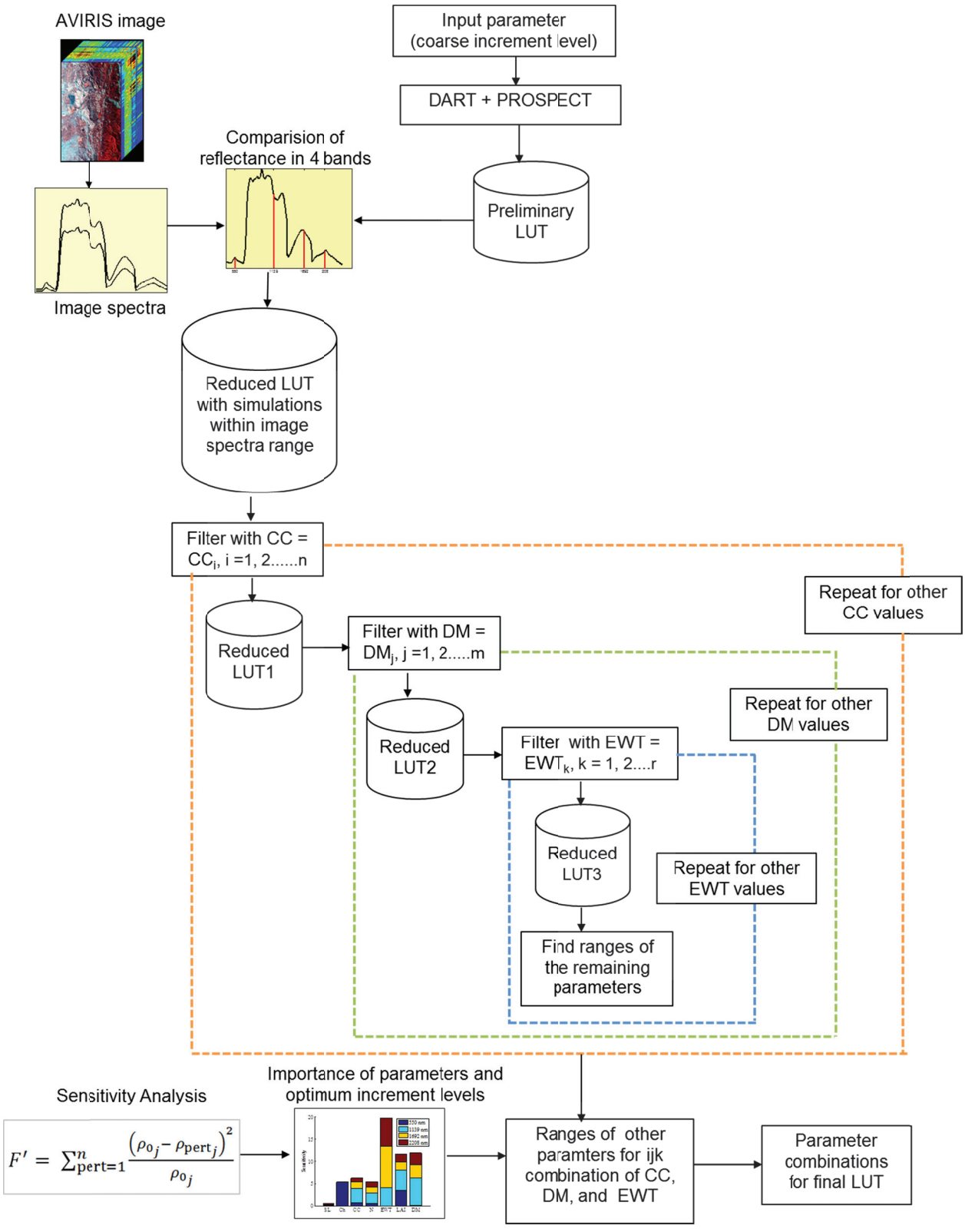

Figure 5: Conceptual framework for the LUT building technique used in this study. In the figure, CC refers to canopy cover (%), DM refers to leaf dry matter content (g/cm^{-2}), and EWT refers to leaf equivalent water thickness (cm).

Initially, a preliminary LUT with a wide range and coarse increments for parameters was constructed for only four narrow AVIRIS bands (550 nm, 1139 nm, 1692 nm, and 2208 nm). PROSPECT (Jacquemoud et al., 1993), which is built into the DART interface, was used to calculate leaf optical properties (reflectance and transmittance). PROSPECT calculates leaf hemispherical transmittance and reflectance as a function of four input parameters, leaf structural parameter (N), leaf chlorophyll a + b concentration (Cab, $\mu g/cm^{-2}$), leaf dry matter content (DM, g/cm^{-2}), and equivalent water thickness (EWT, cm). Together with the PROSPECT parameters, the parameters that were systematically varied (free parameters) were leaf area index (LAI), leaf angle distribution (LAD), soil reflectance (SL), and crown cover (CC). The range for some of these parameters (Cab, LAI, DM, and EWT) were set based upon the knowledge derived from field measurements.

The simulations in the LUT were then compared with the image reflectance to find the optimal ranges for individual parameters. Realistic combinations of parameter values that produced similar output to image reflectance were determined as the range for the rest of model parameters for different combinations of CC, DM, and EWT. Once the optimal ranges and realistic combinations were determined, a simple sensitivity analysis (SA) was performed as in Santis et al. (2009) to determine the suitable incremental steps of parameter values. Based upon the results, a final LUT with finer increments for each important parameter within their narrowed ranges was built. Table 1 shows the free parameters, their minimum and maximum values, and incremental steps used for building the final LUT. The three leaf angle distributions not shown in the table were planophile, plagiophile, and erectophile LADs.

If we had considered all the parameter combinations, 134,400 (16 * 10 * 7 * 2 *5 *4*3) simulations for each AVIRIS band would have been required. After searching for realistic parameter combinations, only 59,60 simulations were needed in the infrared bands and 1,920 in the visible bands. Simulations were done separately for visible and infrared bands in order to reduce computation time, as reflectances in the two regions were sensitive to different set of PROSPECT parameters. Common parameter values in both parts of the spectrum were later used to append the two tables.

Table 1: Parameters, their ranges and increments for the final simulation. The final column shows the total number of values used for each parameter. The parameters are leaf area index (LAI), leaf equivalent water thickness (EWT), leaf dry matter content (DM), leaf structural parameter (N), leaf chlorophyll a + b concentration (Cab), canopy cover (CC) and leaf angle distribution (LAD). The three LADs used were planophile, plagiophile, and erectophile.

Parameters	Minimum	Maximum	Increment	Count
LAI (m^2/m^2)	2.75	6.75	0.25	16
EWT (cm)	0.003	0.0183	0.0017	10
DM (g/cm^{-2})	0.001	0.0132	0.0017	7
N	1.75	2.25	0.5	2
Cab ($\mu g/cm^{-2}$)	20	60	10	5
CC (%)	70%	10%	100%	4
LAD	-	-	-	3

3.6 Calculation of discrete wavelet coefficients

When computing the DWT, two input parameters are required: the choice of mother wavelet and the level of decomposition. We chose the Haar mother wavelet as recent investigations have illustrated its effectiveness in hyperspectral data analysis (Banskota et al., 2011; Bruce et al., 2001; Li, 2004; Zhang et al., 2006). The decomposition level is determined by the type of wavelet and the original signal length. In this study, the level was chosen such that it is maximized (6 for 184 bands using the Haar wavelet). The

DWT was implemented in Matlab (version 7.4, The Mathworks, Inc.) using a dyadic filter tree as described in Section 2. The hyperspectral signal in the spectral domain extracted for each pixel location was passed through a series of low pass and high pass filters. All the detail wavelet coefficients calculated at each level and approximation coefficients at final level were concatenated to produce a test dataset with 17 cases. We also calculated similar coefficients for each simulated reflectance in LUT and concatenated them with the corresponding model parameters to construct a wavelet decomposed LUT database.

3.7 LUT inversion

LUT inversion required similarity matching between wavelet coefficients from plot spectra (measured) and the wavelet coefficients from simulated spectra (modeled). Similarity was assessed using a least root mean square error (RMSE) comparison of the simulated and measured coefficients according to Eq. 3:

$$\mathrm{RMSE}_r = \sqrt{\frac{\sum_{i=1}^{n}\left(W_{measured_i} - W_{modeled_i}\right)^2}{n}} \quad (3)$$

where, $W_{measured}$ is a wavelet coefficient of measured reflectance in test dataset and $W_{modeled}$ is a wavelet coefficient of modeled reflectance in the LUT, and n is the number of coefficients. The solution of the inverse problem was the set of input parameters corresponding to the reflectance in the database that provided the smallest RMSE_r. Because of potential insufficiency in model formulation and calibration and preprocessing errors in measured reflectance, the least RMSE_r solution might not necessarily provide the best estimates of LAI. According to Weiss et al. (2000), the best

parameter retrieval is achieved when the number of solutions q ranges between 10 and 50. Consequently, for each of the 17 cases considered in the test dataset, q = {10, 20, 30, 40 and 50} closest matching LUT entries were selected according to Eq. 3. Out of the multiple solutions, we selected the median LAI value as a final solution.

In DWT, a large number of coefficients usually contain very low values (near-zero), and do not retain any useful information. For example, only 42 out of total 184 wavelet coefficients had 99.99% of the total energy (equivalent to variance) corresponding to original reflectance in one of the plots. The rest of the coefficients (142) had very low values (cumulatively containing only 0.01%of total energy) and might cause inversion bias. Hence, we repeated the inversion procedure but discarded the low coefficients. For each measured wavelet spectrum, the amount of retained energy by an individual coefficient was calculated as the square of its value. The wavelet coefficients were then sorted in descending order of their energies. Two sets of coefficients, which together contained 99.99% and 99.0% of the energy of original spectrum, were created. During inversion, similarity measurements between the wavelet coefficients and simulated spectra (Eq. 3) were calculated using only the two subsets. For convenience, we named the wavelet set with all coefficients as ALL COEFFICIENTS and the wavelet set with 99.99% and 99.0% of energy as ENERGY SUBSET 1 and ENERGY SUBSET 2, respectively.

For comparison purposes, LUT inversion was also performed using the original spectral bands (SPECTRAL BANDS) which required assessing the similarity between plot and simulated spectra. As before, this was carried out using a least root mean

square error ($RMSE_r$) comparison, in the case of the measured and modeled spectra, according to Eq. 4:

$$\mathrm{RMSE_r} = \sqrt{\frac{\sum_{i=1}^{n}\left(R_{measured_i} - R_{modeled_i}\right)^2}{n}} \quad (4)$$

where, $R_{measured}$ is a measured reflectance at wavelength λ, $R_{modeled}$ is a modeled reflectance at wavelength λ in the LUT, and *n* is the number of bands.

4. Results

LUT inversions were performed independently for all wavelet coefficients (ALL COEFFICIENTS), high energy wavelet coefficients (ENERGY SUBSET 1 and ENERGY SUBSET 2), and untransformed original spectral reflectance (SPECTRAL BANDS). All three sets of LUT were sorted from minimum to maximum according to the cost function (RMSE).

The final results are summarized in Table 2. The first column of the table shows the number of solutions considered. The other columns show RMSE and R^2 between model inverted LAI and field measured LAI for ALL COEFFICIENTS, ENERGY SUBSET 1, ENERGY SUBSET 2, and SPECTRAL BANDS respectively.

Table 2: Results from LUT inversion. The first column shows the total number of solutions selected by least RMSEr to calculate the median LAI value. The other columns show the RMSE and R^2 between measured and estimated LAI by inversion using ALL COEFFICIENTS, ENERGY SUBSET 1, ENERGY SUBSET 2, and SPECTRAL BANDS respectively.

Number of solutions (q)	ALL COEFFICIENTS		ENERGY SUBSET 1		ENERGY SUBSET 2		SPECTRAL BANDS	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
10	0.82	0.32	0.62	0.63	0.62	0.44	0.79	0.32
20	0.69	0.43	0.50	0.75	0.60	0.49	0.65	0.46
30	0.69	0.42	0.46	0.77	0.56	0.49	0.60	0.47
40	0.71	0.34	0.49	0.70	0.58	0.41	0.62	0.41
50	0.70	0.35	0.56	0.58	0.58	0.47	0.62	0.40

The results show that the LAI estimated from ENERGY SUBSET 1 provided the greatest accuracy (RMSE = 0.46 and R^2 = 0.77), followed by the LAI estimated from ENERGY SUBSET 2 (RMSE = 0.56 and R^2 = 0.49). Among the different numbers of solutions, the median calculated from 30 solutions provided the best estimate of LAI, the differences were greatest for q = 10 (RMSE = 0.62) and q = 30 (RMSE = 0.46) with ENERGY SUBSET 1. The relations between the measured and model estimated LAI based on the median of 30 solutions from ENERGY SUBSET 1, and SPECTRAL BANDS are shown in Figures 6 and 7. The figures show that LAI is better estimated using ENERGY SUBSET 1 (wavelets) than using SPECTRAL BANDS.

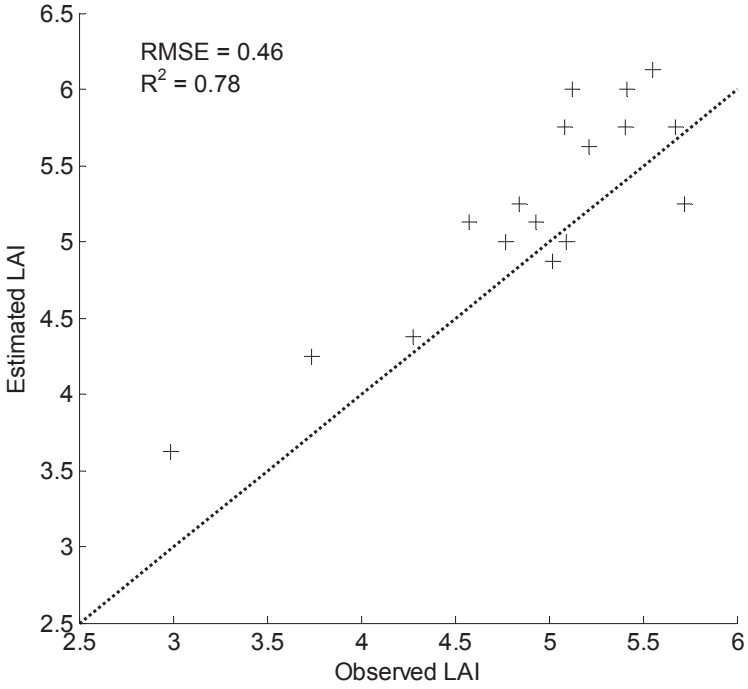

Figure 6: Observed versus best estimated LAI from wavelet coefficients corresponding to 99.99% of total energy (ENERGY SUBSET 1). LAI was estimated as the median value of 30 solutions.

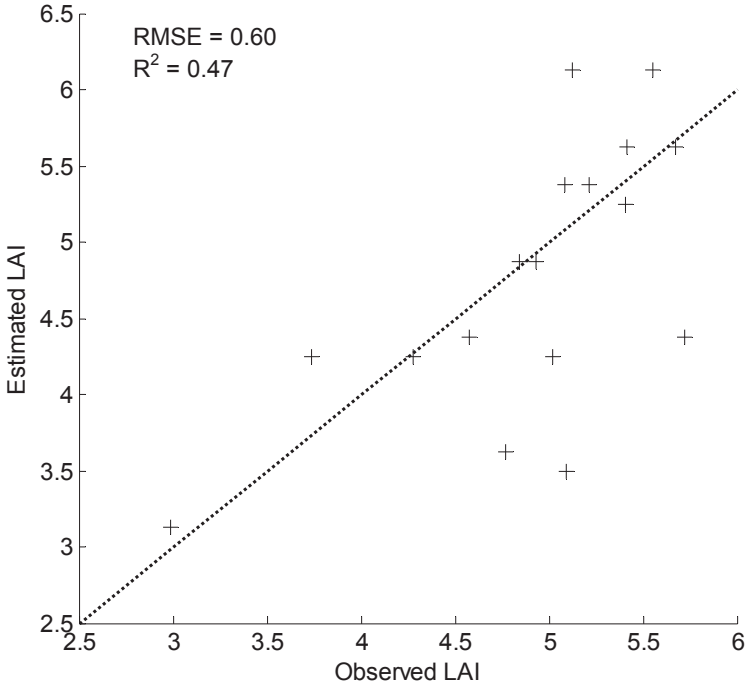

Figure 7: Observed versus best estimated LAI from untransformed AVIRS reflectance (SPECTRAL BANDS). LAI was estimated as the median value of 30 solutions.

5. Discussion

The results indicate that the discrete wavelet transform (DWT) can be a useful tool for estimating forest LAI using hyperspectral data by inverting DART. A simple wavelet decomposition of modeled and measured hyperspectral data by Haar wavelets followed by inversion using high energy wavelet coefficients provided better estimates of LAI than those provided by inversion using the original spectral reflectance.

According to Proposition 3 in Verstraete et al. (1996), in order to invert a reflectance model, the number of collected observations must be greater than the number of the parameters in the model. Hyperspectral data provide dense observational sampling to generate sufficient variance in reflectance measurements and constrain the inversion. However, spectral measurements over narrow bands in hyperspectral data are highly correlated. Such high correlation artificially carries exaggerated weight and causes inversion bias (Bacour et al., 2002; Meroni et al., 2004). The irrelevant bands not sensitive to the model parameters are also more likely to introduce noise instead of useful information and thus adversely affect the inversion accuracy (Schlerf and Atzberger, 2006). Darvishzadeh et al. (2008) did not find any improvement in accuracies by using spectral subsets over all wavebands in estimating grassland LAI from hyperspectral data by inversion of PROSAIL. They selected subsets (1) by stepwise linear regression, (2) based upon the importance of bands for vegetation analysis in the literature, and (3) by choosing bands considered to be well modeled by PROSAIL. On the other hand, Meroni et al. (2004) found that bands selected by a stepwise process in forward modeling performed better than the full spectrum for estimating forest LAI from DIAS airborne hyperspectral data by inverting PROSAIL. Despite its usefulness, the

stepwise procedure for bands selection does not guarantee that the best set of all possible combinations will be found (Meroni et al., 2004; Weiss et al., 2000). Such procedure also necessitates multiple forward runs which are not suitable for complex models like DART.

The results of this study show that the wavelet transformation of hyperspectral bands can be a useful alternative to the band selection techniques employed by previous studies. Wavelet features in this study were less correlated than the original reflectance bands owing to the orthogonal nature of the transformation. DWT also afforded dimensionality reduction, as 99.99% of the total energy was concentrated in few coefficients. Moreover, DWT provided a multi-resolution representation of the original signal in which the effects of different parameters on canopy reflectance can possibly be distributed over wavelet coefficients of different scales. A narrow band in the red edge region of the reflectance spectrum might carry signatures of different parameters including leaf pigments, LAI, etc. After wavelet decomposition, the effect of leaf pigments might be relegated to a small-scale wavelet coefficient, whereas the effect of LAI might be more pronounced in one of the large scale wavelet coefficients (as LAI generally tends to affect reflectance over a larger part of the spectrum than leaf pigments). Such decoupling of the effects over wavelet coefficients at multiple scales and consequently greater independence among parameters is a useful property for inversion. The greatest accuracy obtained by the LUT inversion using ENERGY SUBSET 1 might be attributed to one or all of these benefits offered by DWT.

The poor estimates of LAI by ALL COEFFICIENTS compared to ENERGY SUBSET 1 and ENERGY SUBSET 2 demonstrate how essential it is to select optimal wavelet

features. The selection of wavelet features employed in this study was relatively straightforward as only a small fraction of coefficients retained significant information (as measured by energy). The ENERGY SUBSET 1 discarded the low coefficients and only included the top coefficients cumulatively adding up to 99.99% of total energy. By doing so, the latter represented the original reflectance with a limited number of coefficients without causing significant errors for signal representations (loss of 0.01% of total energy). One of the important properties of DWT is that the energy of the Gaussian noise component of the signal will usually be dispersed as relatively small coefficients (Peng et al., 2009). Hence, getting rid of the low coefficients might also have helped us eliminate the noise.

To understand the effect of different thresholds, we repeated the analysis using a new threshold of 99.0% energy. This led to a decrease in the best accuracy of LAI estimates from RMSE = 0.4638 to RMSE = 0.5613. With 99.0% energy, only 13 coarse-scale coefficients were retained (three approximations and others from scale level 4, 5 and 6), whereas, the 99.99% threshold level retained both coarse scale and fine scale detail coefficients from level 1 to 6. These coefficients from different levels are functions of scale and position (fine detail versus global behavior at various locations in the hyperspectral signal). The greater accuracy by ENERGY SUBSET 1, which comprises both coarse- and fine-scale coefficients, than the ENERGY SUBSET 2 implies that both the detail information and the overall trends of the hyperspectral signals are important for accurate parameter estimations by model inversion.

Figures 8 and 9 show reconstruction of reflectance spectrum for one of the plots using wavelet coefficients belonging to the 99.99% and 99.0% energy thresholds. The

reconstructed curve was more similar to the original reflectance curve in Figure 8 than in Figure 9. Since the 99.0% threshold level selected only coarse level wavelet coefficients, the overall size and shape of the curve was preserved, but some of the important details in visible and near-infrared parts were missing. It might be possible that the best threshold level lies somewhere between 99.0% and 100% other than 99.99% used in this study, but we did not undertake a detailed analysis of the best threshold level since the main objective of this study was to demonstrate the effectiveness of DWT in DART model inversion.

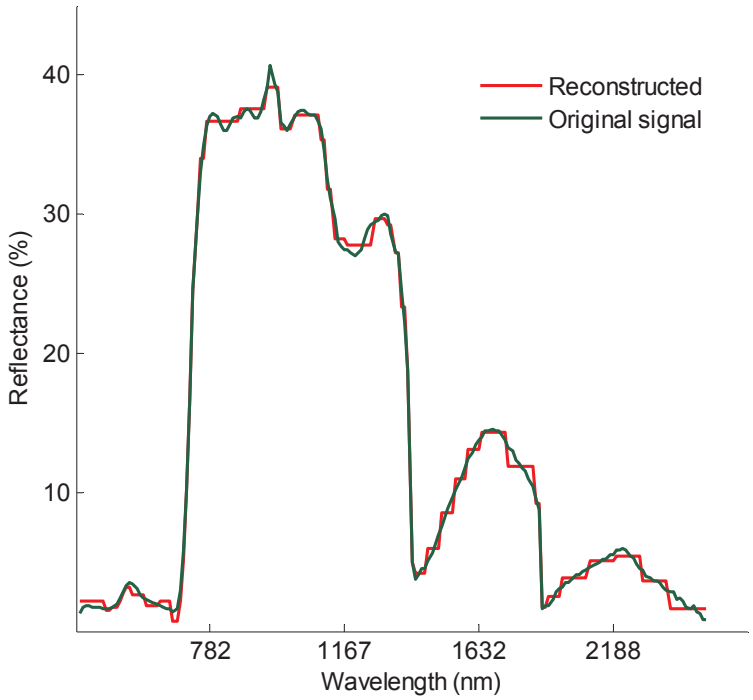

Figure 8: Reconstructed (with coefficients having cumulative 99.99% energy) versus original signal for one of the plots

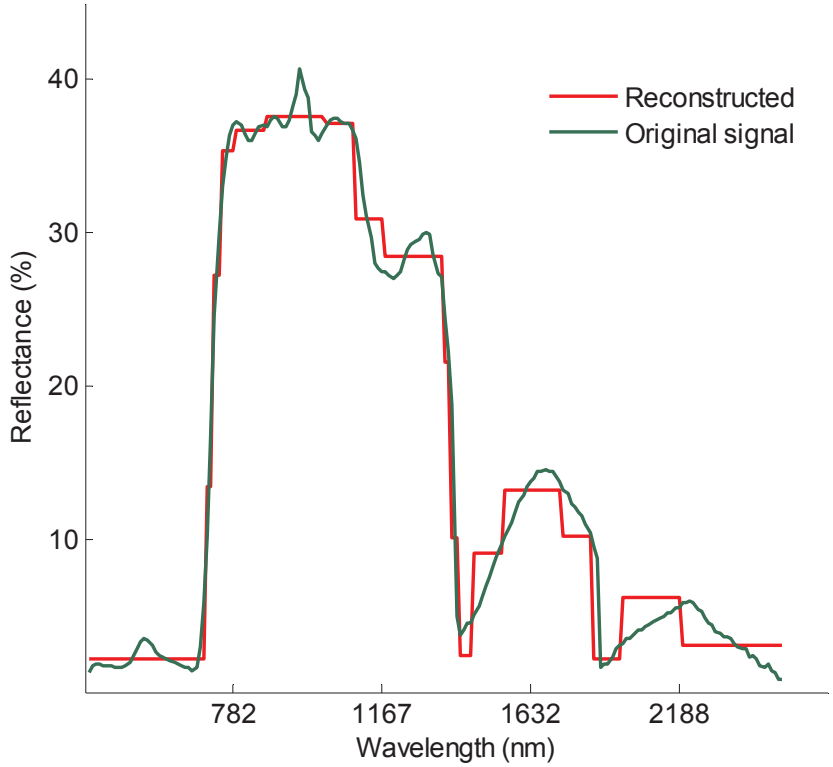

Figure 9: Reconstructed (with coefficients having cumulative 99.0% energy) versus original signal for one of the plots.

The results also demonstrate that the different number of solutions estimated LAI with different accuracies. The differences were greatest between q = 10 and q = 30 with the latter providing the best estimates for all three datasets. This is consistent with the results obtained by Weiss et al. (2000) as they also found $20 < q < 50$ as the optimum number of solutions. RMSE increased when q increased above 30 because the distribution of solutions in terms of LAI was too wide to obtain a good estimate (median value). The highest RMSE with q = 10 indicated that the solution closely matching the test dataset might not be the best solution to estimate LAI.

6. Conclusion

The main objective of this study was to investigate the utility of DWT to estimate forest LAI via DART model inversion using airborne hyperspectral data (AVIRIS). Three sets of wavelet features were examined for their effect on the accuracy of the model inversion. Our results suggest that the Haar discrete wavelet transform can be an effective tool for accurate DART model inversion. We chose to use the Haar wavelet rather than any other family of wavelets because of its simplicity and its usage by previous studies using hyperspectral data. The theory of wavelet transforms is still developing, and more wavelet families are being introduced. The exploration of the utility of other wavelets in model inversion is well warranted since different families of wavelets are suited to different signals and applications.

References

Bacour, C., Jacquemoud, S., Leroy, M., Hautecoeur, O., Weiss, M., Prevot, L., Bruguier, N., & Chauki, H. (2002). Reliability of the estimation of vegetation characteristics by inversion of three canopy reflectance models on airborne POLDER data. *Agronomie*, 22, 555-565.

Baldocchi, D. (2008). Breathing of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. *Australian Journal of Botany, 56*, 1-26.

Banskota, A., Wynne, R.H., & Kayastha, N. (2011). Improving within-genus tree species discrimination using the discrete wavelet transform applied to airborne hyperspectral data. *International Journal of Remote Sensing*, 32, 3551-3563.

Barnsley, M.J., Lewis, P., O'Dwyer, S., Disney, M.I., Hobson, P., Cutter, M., & Lobb, D. (2000). On the potential of CHRIS/PROBA for estimating vegetation canopy properties from space. *Remote Sensing Reviews*, 19, 171-189.

Blackburn, G.A., & Ferwerda, J.G. (2008). Retrieval of chlorophyll concentration from leaf reflectance spectra using wavelet analysis. *Remote Sensing of Environment,* 112, 1614-1632.

Blackburn, G.A. (2007). Wavelet decomposition of hyperspectral data: a novel approach to quantifying pigment concentrations in vegetation. *International Journal of Remote Sensing,* 28*,* 2831–2855.

Bradshaw, G.A., & Spies, T.A. (1992). Characterizing canopy gap structure in forests using wavelet analysis. *Journal of Ecology,* 80, 205-215.

Bruce, L.M., Morgan, C., & Larsen, S. (2001). Automated detection of subpixel hyperspectral targets with continuous and discrete wavelet transforms. *IEEE Transactions on Geoscience and Remote Sensing,* 39, 2217-2226.

Chen, J.M., Govind, A., Sonnentag, O., Zhang, Y.Q., Barr, A., & Amiro, B. (2006). Leaf area index measurements at Fluxnet-Canada forest sites. *Agricultural and Forest Meteorology, 140*, 257-268.

Chopping, M., Moisen, G.G., Su, L., Laliberte, A., Rango, A., Martonchik, J.V., & Peters, D.P.C. (2008). Large area mapping of southwestern forest crown cover, canopy height, and biomass using the NASA Multiangle Imaging Spectro-Radiometer. *Remote Sensing of Environment*, 112, 2051-2063.

Cook, B.D., Bolstad, P.V., Naesset, E., Anderson, R.S., Garrigues, S., Morisette, J.T., Nickeson, J., & Davis, K.J. (2009). Using LiDAR and Quickbird data to model plant production and quantify uncertainties associated with wetland detection and land cover generalizations. *Remote Sensing of Environment, 113*, 2366-2379.

Cook, R.D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, 74, 169-174.

Curran, P.J. (1994). Imaging spectrometry. *Progress in Physical Geography*, 18, 247-266.

Curtis, J.T. (1959). *The vegetation of Wisconsin: An ordination of plant communities*. (2nd ed.). Madison: The University of Wisconsin Press.

Darvishzadeh, R., Skidmore, A., Schlerf, M., & Atzberger, C. (2008). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in heterogeneous grassland. *Remote Sensing of Environment*, 112, 2592-2604.

Desai, A.R., Noormets, A., Bolstad, P.V., Chen, J.Q., Cook, B.D., Davis, K.J., Euskirchen, E.S., Gough, C., Martin, J.G., Ricciuto, D.M., Schmid, H.P., Tang, J.W., & Wang, W.G. (2008). Influence of vegetation and seasonal forcing on carbon dioxide fluxes across the Upper Midwest, USA: Implications for regional scaling. *Agricultural and Forest Meteorology, 148*, 288-308.

Duthoit, S., Demarez, V., Gastellu-Etchegorry, J.P., Martin, E., & Roujean, J.L. (2008). Assessing the effects of the clumping phenomenon on BRDF of a maize crop based on 3D numerical scenes using DART model. *Agricultural and Forest Meteorology*, 148, 1341-1352.

Fernandes, R.A., Miller, J.R., Chen, J.M., & Rubinstein, I.G. (2004). Evaluating image-based estimates of leaf area index in boreal conifer stands over a range of scales using high-resolution CASI imagery. *Remote Sensing of Environment*, 89, 200-216

Gascon, F., Gastellu-Etchegorry, J.P., Lefevre-Fonollosa, M.J., & Dufrene, E. (2004) . Retrieval of forest biophysical variables by inverting a 3-D radiative transfer model and using high and very high resolution imagery. *International Journal of Remote Sensing*, 25, 5601-5616.

Gastellu-Etchegorry, J.P., Demarez, V., Pinel, V., & Zagolski, F. (1996). Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote Sensing of Environment*, 58, 131-156.

Gastellu-Etchegorry, J.P., Gascon, F., & Esteve, P. (2003). An interpolation procedure for generalizing a look-up table inversion method. *Remote Sensing of Environment*, 87, 55-71.

Gastellu-Etchegorry, J.P., Martin, E., & Gascon, F. (2004). DART: a 3D model for simulating satellite images and studying surface radiation budget. *International Journal of Remote Sensing*, 25, 73-96.

Gobron, N., Pinty, B., & Verstraete, M.M. (1997). Theoretical limits to the estimation of the leaf area index on the basis of visible and near-infrared remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 1438-1445.

Goel, N.S. (1988). A perspective on vegetation canopy reflectance models. *In, Proceedings of the Fourth International Colloquium on Spectral Signatures of Objects in Remote Sensing*, Aussois, France.

Gong, P., Pu, R.L., & Miller, J.R. (1995). Coniferous forest leaf-area index estimation along the Oregon transect using compact airborne spectrographic imager data. *Photogrammetric Engineering & Remote Sensing*, 61, 1107-1117.

Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science and Engineering*, 2, 50-61

Green, R.O., Eastwood, M.L., Sarture, C.M., Chrien, T.G., Aronsson, M., Chippendale, B.J., Faust, J.A., Pavri, B.E., Chovit, C.J., Solis, M.S., Olah, M.R., & Williams, O. (1998). Imaging spectroscopy and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, 65, 227-248.

Hedley, J., Roelfsema, C., & Phinn, S.R. (2009). Efficient radiative transfer model inversion for remote sensing applications. *Remote Sensing of Environment*, 113, 2527-2532.

Hsu, P.H. (2007). Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62, 78-92.

Huemmrich, K.F., Privette, J.L., Mukelabai, M., Myneni, R.B., & Knyazikhin, Y. (2005). Time-series validation of MODIS land biophysical products in a Kalahari woodland, Africa. *International Journal of Remote Sensing,* 26, 4381-4398.

Jacquemoud, M. (1993). Inversion of the PROSPECT + SAIL Canopy reflectance model from AVIRIS equivalent spectra: theoretical study. *Remote Sensing of Environment*, 44, 281-292.

Ji, C., Wald, D.J., & Helmberger, D.V. (2002). Source description of the 1999 Hector Mine, California, Earthquake, part I: Wavelet domain inversion theory and resolution analysis. *Bulletin of the Seismological Society of America*, 2002, 92, 1192-1207.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., & Baret, F. (2004). Review of methods for in situ leaf area index determination - Part i. theories, sensors and hemispherical photography. *Agricultural and Forest Meteorology*, 121, 19-35.

Kaewpijit, S., Le Moigne, J., & El-Ghazawi, T. (2003). Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 863–871.

Kimes, D., Knjazikhin, Y., Privette, J., Abuelgasim, A., & Gao, F. (2000). Inversion methods for physically based models. *Remote Sensing Reviews*, 18.

Kimes, D.S., Nelson, R.F., Manry, M.T. & Fung, A.K. (1998). Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *International Journal of Remote Sensing*, 19, 2639–266.

Knyazikhin, Y., Martonchik, J., Myneni, R., Diner, D., & Running, S. (1998). Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data, *Journal of Geophysical Research*, 103, 32257–32276.

Koetz, B., Baret, F., Poilvé, H., & Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy characteristics. *Remote Sensing of Environment*, 95,115−124.

Li, J. (2004). Wavelet-based feature extraction for improved endmember abundance estimation in linear unmixing of hyperspectral signals. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 644 – 649.

Liang, S. (2007). Recent developments in estimating land surface biogeophysical variables from optical remote sensing. *Progress in Physical Geography*, 31, 501-516.

Leblanc, S.G., & Chen, J.M. (2001). A practical scheme for correcting multiple scattering effects on optical LAI measurements. *Agricultural and Forest Meteorology*, 110, 125-139.

Leblanc, S.G., Chen, J.M., Fernandes, R., Deering, D.W., & Conley, A. (2005). Methodology comparison for canopy structure parameters extraction from digital hemispherical photography in boreal forests. *Agricultural and Forest Meteorology*, 129, 187-207.

Malenovski, Z., Martin, E., Homolová, L., Gastellu-Etchegorry, J.P., Zurita-Milla, R., Schaepman, M.E., Pokorný, R., Clevers, J.G.P.W., & Cudlín, P. (2008). Influence of woody elements of a Norway spruce canopy on nadir reflectance simulated by the DART model at very high spatial resolution. *Remote Sensing of Environment*, 112, 1-18.

Mallat, S. (1989). A theory for multi-resolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 674–693.

Martin, M.E., & Aber, J.D. (1997). High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecological Applications*, 7, 431-443.

Meroni, M., Colombo, R., & Panigada, C. (2004). Inversion of a radiative transfer model with hyperspectral observations for LAI mapping in poplar plantations. *Remote Sensing of Environment*, 92, 195-206.

Miller, E.L., & Willsky, A.S. (1996). Wavelet-based methods for the nonlinear inverse scattering problem using the extended Born approximation. *Radio Science*, 31, 51-65.

Mobley, C.D., Sundman, L.K., Davis, C.O., Bowles, J.H., Downes, T.V., Leathers, R.A., Montes, M.J., Bissett, W.P., Kohler, D.D.R., Reid, R.P., Louchard, E.M. & Gleason, A. (2005). Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables, *Applied Optics*, 44, 3576–3592.

Moran, J.M., & Hopkins, E.J. (2002). *Wisconsin's weather and climate*. Madison: The University of Wisconsin Press.

Morisette, J.T., Baret, F., Privette, J.L., Myneni, R.B., Nickeson, J.E., Garrigues, S., Shabanov, N.V., Weiss, M., Fernandes, R.A., Leblanc, S.G., Kalacska, M., Sanchez-Azofeifa, G.A., Chubey, M., Rivard, B., Stenberg, P., Rautiainen, M., Voipio, P., Manninen, T., Pilant, A.N., Lewis, T.E., Iiames, J.S., Colombo, R., Meroni, M., Busetto, L., Cohen, W.B., Turner, D.P., Warner, E.D., Petersen, G.W., Seufert, G., & Cook, R. (2006). Validation of global moderate-resolution LAI products: a framework proposed within the CEOS land product validation subgroup. *IEEE Transactions on Geoscience and Remote Sensing*, 44, 1804-1817.

Myneni, R. B. (1991). Modeling radiative transfer and photosynthesis in three dimensional vegetation canopies, *Agricultural and Forest Meteorology*, 55, 323–344.

Myneni, R.B., Asrar, G., Tanre, D., & Choudhury, B.J. (1992). Remote-sensing of solar-radiation absorbed and reflected by vegetated land surfaces. *IEEE Transactions on Geoscience and Remote Sensing*, 30, 302-314.

Pastor, J., Aber, J.D., McClaugherty, C.A., & Melillo, J.M. (1984). Above-ground production and N and P cycling along a nitrogen mineralization gradient on blackhawk island, Wisconsin. *Ecology*, 65, 256-268.

Pinty, B., Widlowski, J.L., Taberner, M., Gobron, N., Verstraete, M.M., Disney, M., Gascon, F., Gastellu, J.P., Jiang, L., Kuusk, A., Lewis, P., Li, X., Ni-Meister, W., Nilson, T., North, P., Qin, W., Su, L., Tang, S., Thompson, R., Verhoef, W., Wang, H., Wang, J., Yan, G., & Zang, H. (2004). Radiation transfer model intercomparison (RAMI) exercise: Results from the second phase. *Journal of Geophysical Research-Atmospheres*, 109, 19.

Peng, Z.K., Jackson, M.R., Rongong, J.A., Dhu, F.L., & Parkin, R.M. (2009). On the energy leakage of discrete wavelet transform. *A Mechanical Systems and Signal Processing,* 23, 330– 343.

Pu, R., & Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, 91, 212-224.

Santis, A.D., Chuvieco, E., & Vaughan, P.J. (2009). Short-term assessment of burn severity using the inversion of PROSPECT and GeoSail models. *Remote Sensing of Environment*, 113, 126-136.

Schlerf, M., & Atzberger, C. (2006). Inversion of a forest reflectance model to estimate structural canopy variables from hyperspectral remote sensing data. *Remote Sensing of Environment*, 100, 281-294.

Schaepman, M.E., Koetz, B., Schaepman,S.G., & Itten, K.I. (2005). Spectrodirectional remote sensing for the improved estimation of biophysical and -chemical variables: two case studies. *International Journal of Applied Earth Observation and Geoinformation*, 6, 271-282.

Schaepman, M.E., Ustin, S.L., Plaza, A.J., Painter, T.H., Verrelst, J., & Liang, S. (2009). Earth system science related imaging spectroscopy-An assessment. *Remote Sensing of Environment*, 113, S123-S137.

Verstraete, M.M., Pinty, B., & Myneni, R.B. (1996). Potential and limitations of information extraction on the terrestrial biosphere from satellite remote sensing. *Remote Sensing of Environment*, 58, 201-214.

Weiss, M., Baret, F., Myneni, R.B., Pragnere, A., & Knyazikhin, Y. (2000). Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie*, 20, 3-22.

Wessman, C.A., Aber, J.D., Peterson, D.L., & Melillo, J.M. (1988). Remote-sensing of canopy chemistry and nitrogen cycling in temperate forest ecosystems. *Nature*, 335, 154-156.

Zhang, Y.Q., Chen, J.M., & Miller, J.R. (2005). Determining digital hemispherical photograph exposure for leaf area index estimation. *Agricultural and Forest Meteorology*, 133, 166-181.

Zhang, J., Rivard, B., Sanchez-Azofeifa, A., & Casto-Esau, K. (2006). Intra- and inter-class spectral variability of tropical tree species at La Selva, Costa Rica: Implications for species identification using HYDICE imagery. *Remote Sensing of Environment,* 105, 129-141.

CHAPTER 6

CONCLUSIONS

In comparison to multispectral remote sensing, the availability of a large number of narrow spectral bands provides hyperspectral data with important competitive advantages for different vegetation related applications. The capability to analyze canopies using absorption features and over a near continuous spectrum with hyperspectral data has allowed significantly improved species classification and empirically- and physically-based (inversion) modeling of LAI. However, the need for a suitable dimensionality reduction technique has remained a critical challenge for many applications when using hyperspectral data. Discrete wavelet transforms (DWT), through multiscale analysis, offer oppurtunities to both reduce dimensionality and to convey information related to narrow absorption features and the reflectance continuum. To our knowledge, no previous study has evaluated the DWT for classifying species with similar spectral characteristics (e.g., species-level data within the same genus), and modeling LAI (or any other related canopy variables). A previous study by Pu and Gong (2004) [a] tested the dimensionality reduction property of DWT for empirical modeling of LAI, but it did not fully investigate the utility of DWT for multiscale analysis of hyperspectral data.

[a]Pu, R., & Gong, P. (2004). Wavelet transform applied to EO-1 hyperspectral data for forest LAI and crown closure mapping. *Remote Sensing of Environment*, 91, 212-224.

This study showed that Haar DWT features can discriminate pine species with better accuracy than the original AVIRIS spectral bands. Stepwise discriminant analysis selected wavelet detail coefficients related to different levels of decomposition (or scales) and wavelength positions. The detail coefficients that afforded better separation among classes were from the red edge and the visible regions, as well as coefficients concentrated around water absorption features. These wavelength regions have been found to be important for species discrimination by previous studies. Selection of the coefficients from different levels (or scales) suggests that spectral information at both narrow and broad spectral scales is important. The fine scale coefficients were related to subtle changes in band reflectances around the red edge and water absorption regions in the near-infrared (NIR) bands, whereas large-scale coefficients were related to gradual spectral variation around the red-edge, and NIR regions. The lowest separability among pine species was observed when wavelet energy feature vectors, calculated as the sum of square of coefficients at each scale, were used as inputs to the classification, confirming the importance of both fine- and coarse-scale detail coefficients when discriminating among pine species.

For empirical models of LAI, subsets of Haar coefficients from AVIRIS hyperspectral data provided better prediction accuracy than subsets of spectral bands in both deciduous and mixed vegetation types. The subsets were chosen with a genetic algorithm, using multiple linear regressions. The best subset chosen for deciduous plots contained three fine-level (1-level and 2-level) detail coefficients related to the red-edge and NIR (arround 1300 nm) wavelength regions. These wavelength regions have been found to be greatly sensitive to variation in LAI by previous studies. The subset also

chose one coarse scale coefficient related to the visible and red-edge regions. The best accuracy provided by the wavelet subset indicates the multiscale sensitivity of the reflectance to variation in LAI and highlights the importance of DWT for capturing such variation. On the other hand, the best subset for combined vegetation types selected only the fine scale coefficients related to red-edge, NIR, and SWIR reflectances. As such, the selected fine scale coefficients might potentially be less sensitive to background variation caused by different vegetation types and more sensitive to the variation caused by LAI than the large-scale coefficients.

A complex physically-based model like DART can provide a realistic description of the forest canopy reflected radiation. However, it requires an unreasonably high computation time for DART to build a LUT of sufficient size and resolution and to ensure a high degree of accuracy for the estimated LAI by model inversion. In this study, a three-stage approach for building a LUT was tested. The approach involved building a preliminary LUT using model parameters with coarse increments and wide ranges followed by comparisons between LUT and image reflectances. Such comparisons identified parameter combinations close to those occurring in nature at the time and location of image acquisition. A sensitivity study further determined the optimal number of cases for parameters. A final LUT was built based upon the knowledge derived from the preliminary LUT and from the sensitivity study. The results demonstrated that the three-stage approach can greatly reduce the total simulation time for LUT reflectances and hence provide improved opportunities for accurate LAI estimation by DART inversion using hyperspectral data.

For physically-based LAI estimation using DWT, measured reflectances from AVIRIS data and modeled reflectances in LUT were transformed into the Haar wavelet coefficients. Three different subsets of wavelet coefficients were used in the LUT inversion. The first subset contained the full set of wavelet coefficients and the other two subsets contained coefficients that cumulatively added up to 99.0% and 99.99% of the original signal energy. The main reason for testing the two additional subsets with higher energies was to discard a large number of coefficients that have low values (near-zero) and less useful information. The results indicate that the wavelet coefficients from measured hyperspectral data enable better estimates of LAI than subsets using the spectral reflectance bands. The poor accuracy obtained using the wavelet subset with the full set of coefficients suggests that the low-value coefficients represent more noise than signal and should be discarded to avoid inversion bias. The subset with 99.99% energy, which had both fine- and coarse-scale detail coefficients, provided better LAI estimates than the subset with 99.0% energy which had only coarse scale coefficients. This suggests that the information contained in both the narrow and broad parts of the spectrum is important for accurate model inversion. Compared to classification and empirical modeling, the techniques for dimensionality reduction in physically-based modeling are barely explored. To invert a model with a large number of parameters, a sufficient number of measurements (spectral bands) are required, but the measurements should be independent and noise-free. DWT concentrates multiscale information from hyperspectral data in few and less-correlated coefficients and isolates noise components among coefficients with low values. The results of this study showed that when a proper threshold is employed to discard low-value coefficients, the use of

multiscale Haar coefficients offers promise for the accurate estimation of LAI by model inversion.

In a nutshell, this study showed that DWT is a promising tool for hyperspectral data analysis for different vegetation related applications. The properties of DWT that led to improved results in this study can be summarized in four categories. First, the wavelet coefficients reveal multiscale information in narrow and broad wavelength regions, which is not available in the original wavelength domain. Second, a large number of coefficients with low amplitude do not contain useful information, and discarding those coefficients can lead to efficient dimensionality reduction and noise removal. Third, due to the orthogonal nature of the Haar DWT transformation, wavelet coefficients are less correlated than the original hyperspectral bands. Fourth, some fine scale coefficients can normalize the effect of background variation and potentially show increased sensitivity towards the parameters of interest such as LAI and species classes. The best utilization of these properties of DWT, however, requires an optimal method of coefficient selection suited to a particular application. In this study, we used a stepwise approach to select coefficients for pine species classification. Alternatively, use of a genetic algorithm for coefficient selection, as used for empirical estimation of LAI, might result in better classification accuracies. Similarly, a technique for finding the best threshold of energy in LUT-based DART inversion needs further investigation. This study demonstrated the importance of both fine-and coarse-scale coefficients for LAI estimation. Additional information could be obtained by running a sensitivity analysis to analyze the effect of variation in LAI on wavelet features at different positions and scales. Such a study might also help identify features that are more sensitive to other

important biophysical and biochemical parameters of interest and less to background signals caused by soil, crown cover, etc. Extraction of pure sub-pixel signatures for vegetation species (endmembers) is a main requirement for mixed-pixel techniques such as spectral mixture analysis. The suppression of background variation and improved classification of pine species by the Haar coefficients in this study suggest that the DWT can also be a potential tool for mitigating endmember variability in spectral mixture analysis. A study exploring this potential utility of DWT is recommended.

The theory of wavelet transforms is evolving, and more wavelet families are being introduced. Different families of wavelets are suited to different signals and applications. Hence, further studies exploring alternative families of wavelets for related applications might provide better insight into the utility of DWT in hyperspectral data analysis. The significance of such studies might be even greater for data processing from future sensors with higher number of bands and better spectral resolution (e.g., the new generation AVIRIS with a spectral resolution of approximately 0.5 nm).

APPENDIX A: Plot location and associated species in Appomattox Buckingham State Forest
(Data used in Chapter 2)

L = Loblolly Pine, S = Shortleaf Pine and V = Virginia Pine

SN	Longitude	Latitude	Species	SN	Longitude	Latitude	Species	SN	Longitude	Latitude	Species
1	-78.6778	37.4334	L	33	-78.6579	37.4559	L	65	-78.6775	37.4097	L
2	-78.6776	37.4329	L	34	-78.6569	37.4560	L	66	-78.6772	37.4095	L
3	-78.6774	37.4330	L	35	-78.6564	37.4558	L	67	-78.6687	37.4388	L
4	-78.6777	37.4337	L	36	-78.6562	37.4555	L	68	-78.6685	37.4387	L
5	-78.6782	37.4334	L	37	-78.6561	37.4552	L	69	-78.6680	37.4385	L
6	-78.6800	37.4335	L	38	-78.6519	37.4548	L	70	-78.6567	37.4318	L
7	-78.6801	37.4337	L	39	-78.6522	37.4550	L	71	-78.6565	37.4320	L
8	-78.6802	37.4338	L	40	-78.6572	37.4596	L	72	-78.6561	37.4321	L
9	-78.6802	37.4332	L	41	-78.6574	37.4598	L	73	-78.6567	37.4325	L
10	-78.6801	37.4331	L	42	-78.6576	37.4597	L	74	-78.6569	37.4320	L
11	-78.6922	37.4285	L	43	-78.6590	37.4598	L	75	-78.6572	37.4309	L
12	-78.6773	37.4095	L	44	-78.6467	37.4625	L	76	-78.6617	37.4209	L
13	-78.6775	37.4092	L	45	-78.6780	37.4331	L	77	-78.6615	37.4209	L
14	-78.6689	37.4388	L	46	-78.6777	37.4330	L	78	-78.6614	37.4205	L
15	-78.6683	37.4385	L	47	-78.6771	37.4333	L	79	-78.6667	37.4205	L
16	-78.6562	37.4322	L	48	-78.6773	37.4335	L	80	-78.6669	37.4204	L
17	-78.6564	37.4325	L	49	-78.6774	37.4336	L	81	-78.6756	37.4110	L
18	-78.6568	37.4322	L	50	-78.6802	37.4336	L	82	-78.6581	37.4563	L
19	-78.6570	37.4318	L	51	-78.6803	37.4334	L	83	-78.6566	37.4560	L
20	-78.6571	37.4316	L	52	-78.6817	37.4313	L	84	-78.6516	37.4545	L
21	-78.6570	37.4312	L	53	-78.6811	37.4313	L	85	-78.6528	37.4549	L
22	-78.6574	37.4306	L	54	-78.6810	37.4320	L	86	-78.6535	37.4549	L
23	-78.6575	37.4305	L	55	-78.6922	37.4290	L	87	-78.6578	37.4596	L
24	-78.6568	37.4305	L	56	-78.6922	37.4292	L	88	-78.6581	37.4599	L
25	-78.6615	37.4203	L	57	-78.6920	37.4291	L	89	-78.6463	37.4621	L
26	-78.6615	37.4205	L	58	-78.6917	37.4290	L	90	-78.6822	37.4084	S
27	-78.6616	37.4205	L	59	-78.6918	37.4287	L	91	-78.6818	37.4080	S
28	-78.6667	37.4207	L	60	-78.6919	37.4286	L	92	-78.6813	37.4079	S
29	-78.6673	37.4204	L	61	-78.6924	37.4285	L	93	-78.6681	37.4376	S
30	-78.6675	37.4205	L	62	-78.6927	37.4286	L	94	-78.6679	37.4376	S
31	-78.6759	37.4107	L	63	-78.6925	37.4288	L	95	-78.6675	37.4376	S
32	-78.6575	37.4562	L	64	-78.6779	37.4097	L	96	-78.6675	37.4374	S

SN	Longitude	Latitude	Species	SN	Longitude	Latitude	Species	SN	Longitude	Latitude	Species
97	-78.6673	37.43791	S	133	-78.6580	37.4586	S	169	-78.6543	37.4526	V
98	-78.6676	37.43786	S	134	-78.6578	37.4586	S	170	-78.6540	37.4525	V
99	-78.6677	37.43798	S	135	-78.6578	37.4591	S	171	-78.6756	37.4340	V
100	-78.6682	37.4378	S	136	-78.6582	37.4593	S	172	-78.6770	37.4318	V
101	-78.6683	37.4377	S	137	-78.6596	37.4590	S	173	-78.6770	37.4317	V
102	-78.6545	37.4337	S	138	-78.6597	37.4589	S	174	-78.6812	37.4351	V
103	-78.6544	37.43367	S	139	-78.6769	37.4318	V	175	-78.6811	37.4349	V
104	-78.6545	37.43394	S	140	-78.6815	37.4356	V	176	-78.6809	37.4348	V
105	-78.6545	37.45516	S	141	-78.6810	37.4354	V	177	-78.6815	37.4350	V
106	-78.6518	37.45421	S	142	-78.6812	37.4353	V	178	-78.6815	37.4350	V
107	-78.6516	37.4543	S	143	-78.6814	37.4352	V	179	-78.6812	37.4353	V
108	-78.6586	37.45845	S	144	-78.6813	37.4347	V	180	-78.6761	37.4231	V
109	-78.6584	37.45871	S	145	-78.6814	37.4346	V	181	-78.6758	37.4229	V
110	-78.6577	37.45861	S	146	-78.6763	37.4232	V	182	-78.6759	37.4227	V
111	-78.6574	37.45875	S	147	-78.6760	37.4228	V	183	-78.6757	37.4225	V
112	-78.6573	37.45882	S	148	-78.6757	37.4226	V	184	-78.6752	37.4227	V
113	-78.6576	37.45919	S	149	-78.6754	37.4226	V	185	-78.6748	37.4232	V
114	-78.6813	37.40806	S	150	-78.6754	37.4226	V	186	-78.6747	37.4232	V
115	-78.6823	37.40795	S	151	-78.6752	37.4228	V	187	-78.6745	37.4234	V
116	-78.6687	37.43797	S	152	-78.6751	37.4228	V	188	-78.6745	37.4236	V
117	-78.6687	37.43793	S	153	-78.6749	37.4230	V	189	-78.6745	37.4237	V
118	-78.6684	37.4378	S	154	-78.6745	37.4232	V	190	-78.6746	37.4238	V
119	-78.6684	37.43787	S	155	-78.6743	37.4235	V	191	-78.6958	37.4239	V
120	-78.6673	37.43763	S	156	-78.6744	37.4237	V	192	-78.6957	37.4237	V
121	-78.6674	37.43802	S	157	-78.6746	37.4241	V	193	-78.6957	37.4236	V
122	-78.6679	37.43808	S	158	-78.6746	37.4242	V	194	-78.6957	37.4234	V
123	-78.668	37.43808	S	159	-78.6959	37.4239	V	195	-78.6955	37.4236	V
124	-78.6685	37.43795	S	160	-78.6959	37.4238	V	196	-78.6956	37.4238	V
125	-78.6685	37.4382	S	161	-78.6956	37.4235	V	197	-78.6547	37.4549	V
126	-78.6687	37.43843	S	162	-78.6955	37.4237	V	198	-78.6538	37.4537	V
127	-78.6547	37.43332	S	163	-78.6541	37.4540	V	199	-78.6546	37.4536	V
128	-78.6546	37.43338	S	164	-78.6538	37.4539	V	200	-78.6549	37.4539	V
129	-78.6621	37.4214	S	165	-78.6534	37.4534	V	201	-78.6542	37.4528	V
130	-78.6622	37.42128	S	166	-78.6533	37.4533	V	202	-78.6593	37.4654	V
131	-78.652	37.45356	S	167	-78.6546	37.4537	V	203	-78.6599	37.4655	V
132	-78.652	37.45424	S	168	-78.6548	37.4537	V				

APPENDIX B: Wisconsin plot data

(Data used in Chapter 3, chapter 4 and chapter 5)

Site Plot	Longitude	Latitude	Forest Type	Dominant Species	LAI	LAI SD
BH02	-89.9071	43.3724	NH - HEM	TSCA - QURU - QUAL	5.374	0.717
BH03	-89.8454	43.3919	NH	ACSM - TIAM - QURU	5.024	0.399
BH05	-89.8231	43.4080	NH	ACSM	5.409	0.475
BH06	-89.8293	43.4288	OAK-HICK	CAOV - QUAL - QURU	4.842	0.573
BH07	-89.8011	43.4199	NH / MH	POTR - ACSM - QURU	5.121	0.595
BH09	-89.8005	43.3918	NH	ACSM	5.415	1.079
BH10	-89.7619	43.4052	BLH	ACRU - ACSM - PRSE	5.670	0.609
BH11	-89.7465	43.3880	WP	PIST - PODE - ACRU	4.027	0.426
BH12	-89.7401	43.3935	NH - OAK	QURU - QUAL - TIAM	6.743	1.193
BH14	-89.7235	43.4157	OAK - WP	QURU - CAOV - PIST	4.430	0.561
BH15	-89.6977	43.4153	WP - OAK	PIST - QUAL - QURU	4.835	1.056
BH16	-89.7087	43.4352	BLH	FRPE - ACRU - ULAM	3.740	0.469
NC01	-90.7829	45.6242	NH	POTR - ACSM - TIAM	4.933	0.577
NC02	-90.7936	45.6280	NH - HEM	ACSM - TSCA - ACRU	4.159	0.641
NC03	-90.7818	45.6181	NH	ACSM - FRAM - CACO	4.579	0.669
NC04	-90.7753	45.6239	NH	ACSM - FRAM - TIAM	5.210	0.637
NC05	-90.2608	45.9403	MH	ACSM - ABBA - TIAM	5.723	0.993
NC06	-90.2317	45.9432	RP	PIRE - ACRU	4.216	1.808
NC07	-90.2195	45.9758	ASP	POTR	2.989	1.001
NC08	-90.2763	45.9482	BF	LALA - PIMA	2.177	0.358
NC09	-90.2840	45.9714	BLH	FRNI - ABBA - THOC	5.555	0.967
NC10	-90.2257	45.9583	BF	LALA	2.990	0.636
NC11	-90.2293	45.9781	NH	ACSM - TIAM	5.088	0.781
NC12	-90.2509	45.9518	JP	PIBA	3.442	0.783
NC13	-90.2348	45.9615	RP	PIRE - ACRU	4.192	0.897
NC14	-90.3177	45.9657	MX	ABBA - POTR - PIMA	4.244	0.776
NC15	-90.2326	45.9807	NH / BLH	ACSM - TIAM - FRPE	5.095	0.581
NC16	-90.2440	45.9586	JP	PIBA - PIRE - POTR	4.466	0.635
NC17	-90.2591	45.9448	NH	ACSM - FRAM - TIAM	4.767	0.470
NC18	-90.2645	45.9593	WP	PIST	5.253	1.066
NC19	-90.1997	45.9364	RP	PIRE - COAM	3.850	1.053
NC20	-90.2598	45.9796	WP	PIST - ACRU - BEAL	5.615	0.959
NC21	-90.2676	45.9706	MH	POTR - ABBA	4.279	1.215

The first column shows the plot number and site, the second column shows the forest type. The third columns show the name of the dominant species. The descriptions of the dominant and forest type species are provided in Appendix C and Appendix D. Fourth and fifth columns show the mean and standard deviation of plot LAI.

APPENDIX C: Species descriptions

Species Codes	Description
ABBA	Balsam Fir
ACSM	Sugar Maple
ACRU	Red Maple
BEAL	Yellow Birch
CAOV	Shagbark Hickory
CACO	Bitternut Hickory
FRAM	White Ash
FRPE	Green Ash
FRNI	Black Ash
LALA	Tamarack
PIMA	Black Spruce
PIRE	Red Pine
PIST	White Pine
POTR	Trembling Aspen
PRSE	Cherry
QURU	Eastern Red Oak
QUAL	White Oak
THOC	Cedar
TIAM	American Basswood
TSCA	Hemlock

APPENDIX D: Forest type descriptions

Forest type	Description
ASP	Aspen
BF	Boreal Forest (Black Spruce, Tamarack, Lowland Bog)
JP	Jack Pine
NH	Northern Hardwood
NH - HEM	Northern Hardwood, Hemlock
BLH	Bottom Land Hardwood, Wetter Hardwood Site
MH	Mixed Hardwood
MX	Mixed
RP	Red Pine
WP	White Pine